DOCUMENT RESUME

ED 210 748 CS 503 708

TITLE	Perspectives on the Assessment of Speaking and Listening Skills for the 1980s.
INSTITUTION	Northwest Regional Education Lab., Portland, Oreg. Clearinghouse for Applied Performance Testing.
SPONS AGENCY	National Inst. of Education (ED), Washington, D.C.
PUB DATE	Jun 81
GRANT	400-80-0105
NOTE	68p.; Collection of papers prepared for presentation at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).
AVAILABLE FROM	Speech Communication Association, 5105 Backlick Rd., Suite E, Annandale, VA 22003 ($4.25).
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Communication Research; Elementary Secondary Education; *Evaluation Criteria; Evaluation Methods; Higher Education; *Listening Skills; National Surveys; Rating Scales; *Speech Skills; *Test Bias; *Testing Problems

ABSTRACT

The four papers in this collection were drawn from a symposium convened to review the state of the art in speaking and listening assessment at all educational levels and to suggest strategies for addressing the technical issues, particularly those relating to test bias and the use of rating scales, that speaking and listening assessors often face. The first paper presents results of a national survey of elementary and secondary speaking and listening assessment procedures that was sponsored by the Speech Communication Association. Various sections of the paper focus on state legislative activity, specific skills tested, standards, and testing procedures. The paper reports the findings of the survey on a state-by-state basis. The second paper deals with the current assessment of speaking and listening skills at the college level and concentrates on the development and evaluation of the Communication Competence Assessment Instrument. The third paper reviews potential sources of bias in assessment--including stimulus conditions, response mode, rating procedures, and use of test results--and suggests strategies for dealing with each. The fourth paper identifies several issues related to the use of rating scales in evaluating oral communication proficiency--including scale development, psychometric evaluation, and the speech type assessed--and suggests new directions for research. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED210748

Perspectives on the Assessment of Speaking and Listening Skills for the 1980s

Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

June 1981

This collection of papers was prepared for presentation as a symposium at the annual meeting of the American Educational Research Association, Los Angeles, 1981.

CS503708

Available for $4.25 a copy from:
Speech Communication Association
5105 Backlick Road, Suite E
Annandale, VA 22003

This collection of papers was prepared for presentation as a symposium at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

The material presented herein was developed by the Clearinghouse for Applied Performance Testing (CAPT) under grant #400-80-0105 with the National Institute of Education (NIE) of the Department of Education. The opinions expressed in this publication do not necessarily reflect the position of NIE and no official endorsement by the Institute should be inferred.

INTRODUCTION

The assessment of speaking and listening skills is becoming a high priority nationwide. Numerous state level assessment programs consider communication skills important targets for basic skills assessment, and several assessments focus specifically on the assessment of speaking and listening. In addition, the Speech Communication Association has assembled a national task force to document research and development in speaking and listening assessment and to promote sound assessment procedures.

This symposium was convened to provide a relevant, up-to-date picture of current issues, recent advances and needed research relating to the measurement of speaking and listening skills. Specifically, the purposes of this symposium were to (1) review the state of the art in speaking and listening assessment at the elementary, secondary and postsecondary levels, and (2) suggest strategies for addressing technical issues—particularly those relating to test bias and the use of rating scales—faced by speaking and listening assessors.

Speaking and listening assessment is an important topic in educational, in part because it represents an opportunity to use performance assessment, a viable alternative to traditional paper and pencil assessment. Performance assessment relies on the direct observation and evaluation of behavior by qualified judges under real or simulated circumstances. This form of assessment is rapidly gaining acceptance for several reasons:

1. There is growing interest in assessment methods in which results can be generalized to behavior outside school. High fidelity performance tests, such as those designed to measure speaking and listening skills, maximize the external validity of the assessment.

2. Some evidence suggests that performance measures can minimize the adverse impact of testing on minorities. Speaking and listening assessment has a high potential for bias due to the heavy cultural loading of oral communication. However, carefully developed assessment materials and approaches can greatly reduce problems of bias.

3. A growing number of business, vocational and industrial training programs, as well as professional licensing and certification programs provide new settings for the development and use of performance assessment.

The papers included in the symposium present a variety of perspectives on performance assessment in the evaluation of speaking and listening skills. In the first paper, Philip Backlund, of Central Washington University, and his associates present results of a national survey of elementary and secondary speaking and listening assessment procedures sponsored by the Speech Communication Association. The survey focused on state legislative activity, specific skills tested, standards and testing procedures. Results of the survey are reported on a

state-by-state basis. In the second paper, Rebecca Rubin, of the University of Wisconsin-Parkside, deals with the current assessment of speaking and listening skills at the college level. Rubin's paper concentrates on the development and evaluation of the Communication Competence Assessment Instrument.

Richard Stiggins of the Clearinghouse for Applied Performance Testing, Northwest Regional Educational Laboratory, reviews potential sources of bias in speaking and listening assessment and suggests strategies for dealing with each. Specific sources of bias are identified in stimulus conditions, response mode, rating procedures and use of test results. Donald Rubin of the University of Georgia then identifies several important issues related to the use of rating scales in evaluating oral communication proficiency. In particular, Rubin covers scale development, psychometric evaluation and the type of speech assessed, and suggests new directions for research.

The presentation at the Annual Meeting of AERA included a fifth paper, "Referential Communication Activities: A Tool for Assessing Speaking and Listening in the Classroom," by W. Patrick Dickson and Janice Patterson of the University of Wisconsin-Madison. Since the authors will be publishing their study in a research journal, they elected not to include it in this publication.

TABLE OF CONTENTS

	Page
Acknowledgements	i
Symposium Presenters	ii
Introduction	1
A National Survey of State Practices in Speaking and Listening Assessment Philip Backlund, Dwayne VanRheenen, Michael Moore Arlie Muller Parks and James Booth	3
Assessment of College-level Speaking and Listening Skills Rebecca B. Rubin	25
Potential Sources of Bias in Speaking and Listening Assessment Richard J. Stiggins	43
Using Performance Rating Scales in Large-scale Assessments of Oral Communication Proficiency Donald L. Rubin	51

ACKNOWLEDGEMENTS

CAPT would like to express appreciation to authors of the papers included here for granting the Clearinghouse permission to disseminate their thoughts and ideas on speaking and listening assessment. Thanks are also due to Nancy Mead of NAEP for her help in organizing the symposium, and for chairing the AERA convention presentation. Editorial assistance was provided by Nancy Bridgeford and Vicki Spandel. Special thanks to Nancy who, with the able secretarial assistance of Barbara Hejtmanek and Carol DeWitte, coordinated production of this publication.

Richard J. Stiggins
CAPT Coordinator and Symposium Organizer

SYMPOSIUM PRESENTERS

Philip Backlund
Department of Communication
Central Washington University
Ellensburg, WA 98926
(509) 963-1066

Nancy Mead
NAEP
Education Commission of the States
1860 Lincoln Tower
Suite 300
Denver, CO 80295
(303) 830-3742

Donald Rubin
University of Georgia
125 Aderhold Hall
Athens, GA 30602
(404) 542-3030

Rebecca B. Rubin
University of Wisconsin-Parkside
Kenosha, WI 53141
(414) 553-2017

Richard J. Stiggins
Clearinghouse for Applied Performance Testing
Northwest Regional Educational Laboratory
300 S. W. Sixth Avenue
Portland, Oregon 97204
(503) 248-6800, Ext. 352

A NATIONAL SURVEY OF STATE PRACTICES IN SPEAKING AND LISTENING ASSESSMENT

Philip Backlund, Presenter
Central Washington University

Co-authors:
Dwayne VanRheenen
Michael Moore
Arlie Muller Parks
James Booth

ABSTRACT

The Speech Communication Association, through its task force on assessment and testing, has been active in monitoring states' progress in the development of assessment programs in speaking and listening. A subcommittee of the task force was given the task of surveying each state's current practices and plans for the assessment of speaking and listening skills at the elementary and secondary levels. This paper summarizes the survey in three sections. The first section covers the objective of the survey and the survey procedures. The objective was to gather information on legislative action, skills tested, setting skill level standards and testing procedures. The second section summarizes findings. The third section offers some critical comments concerning the findings.

A NATIONAL SURVEY OF STATE PRACTICES IN SPEAKING AND LISTENING SKILL ASSESSMENT

The description of the basic skills in the Elementary and Secondary Education Act (Title II) as "reading and mathematics and effective communication, both written and oral," reflects an expanded view of what is basic in education. The primary area of expansion is oral communication. For decades, research and curriculum materials development in oral communication lagged significantly behind the work done in written communication. This is beginning to change. The expansion of basic skills to include oral communication has led many states and school districts to conduct research and develop instructional materials in oral communication. The Speech Communication Association has attempted and is attempting, through a variety of means, to advise, inform, and coordinate these various efforts.

One such effort is being made through the Task Force on Assessment and Testing. In 1979, the Educational Policies Board of the Speech Communication Association established the Task Force to facilitate the flow of information on assessment of communication skills from members of the speech communication profession to federal, state, and local educational leaders.

The Task Force in turn was organized into three committees. One of these committees is called the "Committee on Current Assessment Practices, K-12." The purposes of this committee are to 1) monitor assessment practices on a state-by-state, and, when appropriate, local district basis; 2) gather and disseminate systematic data describing current practices; and 3) maintain and enlarge the Task Force file on resource personnel and assessment materials. This paper reports the committee's activities and findings as of March 15, 1981.

The report is organized into five sections: A) purposes and procedures, B) the Federal perspective on basic skills and oral communication, C) a summary of the findings, D) conclusions based on data gathered, and E) appendices that present a state-by-state description of activities in oral communication assessment.

Purposes and Procedures

The Committee believes this report serves three major purposes. First, it provides speech communication professionals and others who share their interest with information on state level assessment of oral communication. Second, it provides information on how each state is responding to basic skills legislation regarding speaking and listening. Third, it serves as an information clearinghouse to federal, state, and local school governing agencies concerning oral communication procedures and plans for the governing agencies. The Committee does not view this report as a final product. Laws, regulations, and learning objectives are changing rapidly. The Committee will continue to monitor these activities and report their changes as new information is obtained. In keeping with this, the Committee would appreciate learning of any omissions or inaccuracies the reader may note in this report.

The Committee on Current Practices was created to carry out these purposes. The Committee consists of Chairman Phil Backlund, Central Washington University; Dwayne VanRheenen, University of Maine; Mike Moore, Indiana University; Arlie Muller Parks, Mansfield State College; and James Booth, Murray State University.

To obtain the information, the committee conducted a telephone survey. We talked with either the Basic Skills Coordinator, the Language Arts Coordinator, and/or the Director of Testing and Evaluation in each state. The following questions were asked: 1) How is your state responding to the basic skills legislation in regard to speaking and listening? 2) Does your state have (or is it developing) statewide learning objectives for speaking and listening? 3) Are you currently assessing speaking and/or listening skills, or do you have plans for such an assessment? 4) If so, what instruments are used? 5) Which specific skills are assessed, and at what grade level? 6) How are the results used?

The committee presented its preliminary findings at the annual Speech Communication Association convention in November, 1980. The committee will continue to review and update the report as new data comes in.

An Overview of Federal Activity in Basic Skills

The impetus for much of the interest in basic skills came from federal legislation enacted in 1978. Public Law 95-561 amended the Elementary and Secondary Education Act of 1965 to include (among other items) a definition of basic skills that includes reading, mathematics, oral communication, and written communication. This marked the first time that oral communication had been so included in federal legislation. Title II of this law authorized appropriations to fund the Basic Skills Improvement Program. To provide an example of the type of programs funded under this legislation, here is a summary of the appropriations for fiscal year 1980.

Under Part A of the legislation, three kinds of demonstrations were authorized for the improvement of the basic skills of both school children and youth and adults: In-School (Section 205), Out-of-School (Section 208), and Parent Participation (Sections 206). Total funding for these demonstrations was $15,149,789. In addition, grants and contracts were authorized under Part A for the provision of technical assistance and for dissemination of information about the basic skills. Total cost of these contracts was $2,781,666. Under Part B, grants were authorized for the support of two efforts by state education agencies: under Section 222, states may support demonstration projects in elementary and secondary schools, and under Section 224 they may support staff development in the basic skills at both the state and local levels. Appropriations for these state grants total $7,749,169. Part C of the Basic Skills legislation authorized two special programs--one for improvement of reading through distribution of materials for reading motivation (RIF-Reading is Fundamental) and the other in mathematics. Appropriations for the former were $6,500,000 and for the latter $586,252 (U.S. Dept. of Ed., 1980). The U.S. Department of Education NEWS reported that 156 grants and 15 contracts had been awarded under this program as of January 30, 1981 (NEWS, 1981).

It is important to note, however, that this legislation does not require states to develop programs in each of the basic skills areas. Each state develops its own plans, sets its own priorities, and distributes grant funds as it sees fit. Thus, if a state chooses not to focus on any one of the basic skill areas (such as oral communication), the Basic Skills Office of the Department of Education has no power to require the state to do otherwise. The Basic Skills Improvement Program is set up to assist, not to mandate program content (Broom, 1981).

President Reagan's economic policies have further impact on this program. The President is moving to change many formula grants into block grants. For education this implies that individual states will have further control over their educational programs. Instead of receiving a number of grants for individual programs (such as basic skills), the states will receive a block of money to distribute as they see fit. It is anticipated that this change will be implemented by October 1, 1981. In spite of funding changes, however, the Basic Skills Improvement Program of the U.S. Department of Education expects to provide strong support and assistance services for the individual states in their efforts to develop quality basic skills programs.

Findings Summary

Each of the states surveyed can be placed into one of five categories. The first includes those states that have no curriculum or assessment programs in speaking and/or listening and have no plans of implementing any such programs.

The second category has expressed an intention to develop either curriculum or assessment programs in speaking and/or listening, but have not yet done so. The third category includes states that have identified speaking and/or listening skills and are developing statewide procedures to assess these skills. The fourth category includes states that have identified skills in speaking and/or listening and are developing curriculum materials for teaching these skills but have no assessment procedure. The fifth category includes states that have identified skills and have developed (or are developing) both curriculum and assessment procedures for these skills.

The first category includes nineteen states. These states are Alabama, Alaska, Arkansas, Colorado, Florida, Idaho, Kansas, Kentucky, Maryland, Mississippi, New Mexico, New York, North Carolina, North Dakota, Oklahoma, Tennessee, Washington, West Virginia and Wyoming.

The second category includes fourteen states. These states are California, Connecticut, Delaware, Indiana, Iowa, Maine, Montana, Nebraska, Nevada, New Jersey, Ohio, Oregon, Rhode Island and Wisconsin.

The third category includes four states. These include: Hawaii is in the process of developing a speaking skill rating scale for use statewide in grades 3, 6, 8, 10 and 12. Indiana (Gary Community School District) has developed an oral proficiency examination as part of graduation requirements for its high school students. Massachusetts has identified fourteen speaking and eleven listening skills; testing of these skills has been contracted for with the National Assessment of Educational Progress. New Hampshire assessed a sample of 5th, 9th, and 12th grade students statewide in listening skills in October 1980.

The fourth category includes eleven states. Arizona has developed 122 listening/reading skills and 120 speaking/writing skills broken down by grade level. Georgia has both high school graduation requirements in speaking and listening (local authorities are responsible for testing) and comprehensive Language Arts Guide which details resource and curriculum materials. Illinois is developing a list of specific oral communication skills for use in planning curriculum objectives for K-12. Louisiana has identified twenty-three oral communication skills as part of their writing assessment program. Minnesota has a highly developed statewide curriculum support plan for all the basic skill areas—including speaking and listening. Missouri (Independence School District) has developed specific speaking and listening competencies together with suggested curriculum activities designed to teach these competencies. North Carolina has developed a detailed list of speaking and listening competencies together with performance indicators. South Carolina has developed minimum standards for statewide educational objectives in reading, writing and mathematics. Speaking and listening objectives are a part of reading and writing. South Dakota has published a well developed curriculum guide in speaking and listening to be used by local districts in planning language arts programs. Texas has developed curriculum objectives in speaking and listening for K-12. Virginia is preparing a set of objectives and assessment plans for all basic skills including speaking and listening.

The fifth and final category includes four states. Michigan is developing a program of speaking and listening skills, curriculum strategies to teach those skills, and assessment procedures tied directly to the skills. Pennsylvania is developing a set of classroom assessment instruments and curriculum materials designed to help the classroom teacher identify areas for instructional development in speaking skills. Utah has a well developed statewide program combining basic skills (including speaking and listening) with life skills. Vermont has a statewide program in speaking and listening skill assessment in place and reports statewide test results yearly.

The Speech Communication Association has published detailed state by state summaries. SCA also has copies of many state basic skill plans, curriculum guides, and assessment procedures. Interested readers are urged to contact SCA for further information.

Conclusions

It is not the purpose of this report to critically evaluate any state program. We can, however, make some observations and recommendations concerning state activities in speaking and listening.

1) There is a great deal of interest and activity in oral communication, yet there is little available information. Many state basic skill directors are highly interested in receiving more information on oral communication. Professionals in speech communication and in education can provide a valuable service by contacting their basic skills directors and local school authorities to offer assistance.

2) While many states have programs in oral communication, many do not. These states need to be shown that programs in oral communication are both desirable and feasible.

3) Further coordination between state programs would be useful. We discovered some "reinventing the wheel" in our survey--that is, states developing similar programs without knowing what other states were doing. We can help provide such coordination.

4) It may be useful to urge more consistency in the various definitions of oral communication skills and assessment procedures. Current lack of consistency seems to be causing some coordination problems. It may be useful to urge states to more closely coordinate the identification of these skills, perhaps using the model presented by the Speech Communication Association in its "Standards for Effective Oral Communication Programs."

5) More research is needed in oral communication at the elementary and secondary levels. Specific areas include effectiveness of instructional programs in oral communication, relationship of oral communication skills to written communication skills, and curriculum material development.

In summary, oral communication is receiving much more attention now than only two or three years ago. Many states are including (or planning to include) speaking and listening skills in curriculum and assessment programs at the elementary and secondary levels. Despite significant progress, however, much work must be done before oral communication curricula and assessment methods are as fully developed as the materials in the other basic skill areas. As professionals interested in oral communication, we view that development as our overall goal.

Results of the State Speaking & Listening Survey*

ALASKA: Nothing happening at this time. No plans.

Contact:
Darby Anderson, Director
Right to Read Program
Department of Education
Pouch F - Alaska Office Building
Juneau, AK 99811

ALABAMA: Nothing currently in place or planned for speaking and listening. No definite response to the Basic Skills Legislation as yet.

*Note: This is an abbreviated summary of the original survey.

5616D

ALABAMA: (continued)

Contact:
Mr. Bill Malone
Educational Specialist
State Department of Education
State Office Building, Room 402
Montgomery, AL 36130
(205) 852-3220

ARIZONA: ARS 15-102.26 "Suggested Guidelines for the Development and Implementation of a Continuous Uniform Evaluation System" appears to be their response to Title II. Speaking and listening are not addressed.

Contact:
Mr. Steve Stephens
Research and Statistical Analyst
State Department of Education
1535 West Jefferson
Phoenix, AZ 85007
(602) 255-4888

ARKANSAS: Nothing is happening at this point in time concerning speaking and listening. No apparent plans.

Contact:
Dr. Constance S. Dardin
Coordinator, Student Assessment
Department of Education
ArchFord Education Building
Little Rock, AR 72201
(501) 371-1461

CALIFORNIA: No statewide testing program in speaking and listening. Local school districts have obligations for basic skill improvement programs.

Contact:
Bill Burson
California Department of Education
721 Capitol Mall
Sacramento, CA 95814

COLORADO: Has developed a comprehensive program designed to train local agencies and personnel in identifying their own needs or basic skills and developing corresponding programs. Thus, there are no state guidelines on the curriculum or assessment standards for any of the basic skill areas. Program is called "Colorado Comprehensive State Basic Skills Plan" and was legislated by Public Law 95-561.

COLORADO: (continued)

Contact:
Mr. Robert Ewy, Supervisor
Title II, Basic Skills
Colorado Department of Education
201 E. Colfax
Denver, CO 80203
(303) 839-2232

CONNECTICUT: Recently passed Public Act 128 lists required subjects to be taught in the public schools. Speaking is among those subjects. There is an indefinite plan to assess speaking or listening on a statewide basis in the future.

Contact:
George D. Kinkade, Consultant
Bureau of Research, Planning and Evaluation
State Department of Education
Box 2219
Hartford, CN 06115
(203) 566-7232

DELAWARE: Has a well developed state program titled "Delaware Educational Assessment Programs" initiated in 1972 and currently requiring the annual administration of nationally standardized, norm-referenced achievement tests in reading, language arts, and mathematics. Currently the program uses the California Achievement Test. Speaking and listening are mentioned as "goals", but no plans are currently being made to test these areas.

Contact:
Robert A. Bigelow
State Supervisor of Educational Assessment
State Department of Public Instruction
Townsend Building, Box 1402
Dover, DE 19901
(302) 678-4583

FLORIDA The state has a well developed program for testing certain basic skills as a requirement for graduation, but speaking and listening are not among them. The state basic skills plan mentions oral communication, but does not specify objectives, curriculum materials, or assessment procedures. Speaking and listening currently do not have a high priority.

Contact:
Ms. Ada Puryear
Administrator of Early Childhood & Elementary Education
Florida Department of Education
Knott Building
Tallahassee, FL 32301

GEORGIA:

First, the State Board of Education policy statement 30-770 and 30-710, "High School Graduation Requirements, (Nov. 1976 and amended in 1978 and 1980), which goes into effect with the graduating class of 1984, requires that each student demonstrate at least minimal mastery of the following competency performance standards: reading, writing, mathematics, speaking and listening, and problem solving. For speaking and listening, the requirements read: A student will demonstrate competence by his or her ability to receive and transmit oral and aural communication in the context of academic problems, everyday tasks and employment activities. Indicators of speaking and listening may include interpreting aural communications, composing oral directions and questions and using formal and informal speaking styles.

Local authorities have responsibility for providing evaluation services together with appropriate curriculum.

Second, the State Department of Education is preparing a document called "Comprehensive Language Arts Guide" for distribution to schools. This Guide should be available in June 1981.

Contact:
Dr. William Hammond, Director
Georgia Basic Skills Program
Georgia Department of Education
156 Trinity Avenue
Atlanta, GA 30303
(404) 656-2584

HAWAII:

(A bright spot.) Board of Education Regulation #4540.1, revised 8/78, effective with graduating class of 1983. They have developed specific performance expectations for grades 3, 6, 8, 10 and 12 in language arts, including many related to speaking, listening, and nonverbal communication. A test titled "Oral Communication Test for Essential Competency #4" has been drafted and field tested in April/May 1980. The test currently contains a rating scale similar to the one developed for Massachusetts. Plans call for further development through the 1980/81 school year.

Contact:
Dr. Selvin A. Chin-Chance
Administrator, Test Development & Administration Section
Department of Education
P. O. Box 2360
Honolulu, HI 96814

IDAHO: Proficiency testing is being done in the Idaho public schools on a district optional basis. Approved 2/3/77, it defines basic skills as reading, writing, arithmetic, and spelling. Nothing is included on speaking and listening and there are no current plans.

Contact:
Karen Underwood, Consultant
Proficiency Testing
State of Idaho
Department of Education
Len B. Jordan Office Building
Boise, ID 83720
(208) 384-2112

ILLINOIS: A planning document, prepared by the Illinois Speech and Theater Association in cooperation with the Illinois State Board of Education Department of Program Planning and Development, outlines four major categories of oral communication skills and describes specific communication skills for each. No assessment method is suggested. The four major categories are: 1) providing a wide variety of communication situations; 2) learning to analyze communication messages appropriately; 3) developing appropriate communication strategies; and 4) developing the ability to evaluate messages.

A copy of this plan is in the committee files.

Contact:
Mina Halliday
SBOE Program Planning and Development
State Office of Education
100 North First Street
Springfield, IL 62777

INDIANA: No current statewide efforts until 1983 to either identify speaking and listening skills or to specify assessment procedures. Some local districts have excellent programs in speaking and listening, most notably the Gary Community School Corporation. Commencing with the graduating class of 1980, all students must pass an oral proficiency examination. All middle school students are required to take a minimum of one semester of speech. All high school sophomores are required to take a full semester of speech. An evaluation instrument accompanies the courses: It rates eight areas holistically on a four point scale. The eight skill areas are: articulation, pronunciation, verbal utterances, rate, word usage, voice qualities, volume, and sentence structure.

INDIANA: (continued)

Contact:
Dr. Patrick Gavigan
State Department of Public Education
120 West Market Street
Indianapolis, IN 46204
(317) 633-4355

IOWA: Currently no testing of speaking and listening and no priority to do so. State Department of Public Instruction does have a state plan to train teachers in the assessment of math, reading, and writing, as well as oral communication.

Contact:
Lee Wolf
State Department of Public Instruction
Grimes State Office Building
Des Moines, IA 50319
(515) 281-3190

KANSAS: Nothing at this time in speaking and listening. Tests being developed for reading and mathematics.

Contact:
Mr. Mel Riggs
Education Program Specialist
State Department of Education
120 East 10th Street
Topeka, KS 66612

KENTUCKY: Nothing currently planned on assessing speaking and listening. Has a criterion-referenced test for reading, writing, and arithmetic.

Contact:
Dr. Charles E. Morgan
Director, Testing
State Department of Education
Capitol Plaza Tower
Frankfort, KY 40601

LOUISIANA: In 1977 Louisiana established a program of minimum standards in reading, and later (1979) in writing and mathematics. Speaking and listening are not treated separately; however, each previously mentioned skill area includes some oral communication skills. For example, reading includes, for kindergarten and first grade, skills in auditory discrimination, oral vocabulary, oral expression of ideas, and listening comprehension. The higher the grade, however, the less attention is given to oral communication.

LOUISIANA: (continued)

Contact:
Mr. Joseph Williams
Director, Planning and Evaluation
Bureau of Assessment
State Department of Education
P. O. Box 44064
Baton Rouge, LA 70804

MAINE: By 1981, the State Department will have a plan suggesting models districts can use in developing their own plan. By 1982, each school district must have a written plan for the achievement of basic competencies. Currently the State Department of Education is developing a list of competencies in all areas, including speaking and listening. However, it seems unlikely there will be performance testing, most probably the assessment will be paper and pencil.

Contact:
Dale Douglas
Educational Planner
State of Maine
Department of Educational and Cultural Services
Augusta, MA 04333

MARYLAND: No plans to test speaking and listening. Currently only testing reading.

Contact:
Dr. William Grant, Chief
Program Assessment Branch
State Department of Education
BWI Airport
P. O. Box 8717
Baltimore, MD 21240
(301) 796-8300

MASSACHUSETTS: Has a well developed plan for assessing speaking and listening. The speaking test is currently undergoing refinement. Plans are to have it in place for the eighth grade level by the school year 1981-82; the listening test should be finished by the 1982-83 school year.

Contact:
Tracy Libros
Bureau of Research and Assessment
State Department of Education
31 St. James
Boston, MA 02116
(617) 727-0190

MICHIGAN:

PL#38 called for statewide assessment in language arts, reading, etc. In 1973, criterion-referenced assessment was mandated. The Department of Education developed objectives for the state in 1975-76, including objectives for speaking and listening. In 1977, the state began to assess listening (speaking not yet done, though ready to go). Listening test used on 4th, 7th grades in 1978; on 10th in 1979. Used statewide sample (1,000 - 3,000 per grade level). Test is mandated but is not required for graduation. Eight listening skills (each criterion-referenced) and twelve speaking skills (not yet field tested) are used. Deals with using videotape, pantomiming/roleplaying, public speaking, and small group discussion. State has instructional packages on how to teach and assess these areas.

Contact:
Ed Roeber
Supervisor/MEAP
State Department of Education
620 Michigan National Tower
P. O. Box 30008
Lansing, MI 48909
(517) 373-8393

MINNESOTA:

Minnesota passed a state basic skills law in 1979 that requires school districts to develop programs in five basic skill areas: writing, listening, speaking, reading, and computing. "Essential learning outcomes" for each of the five basic skill areas have been developed. The outcomes for listening and speaking are currently undergoing revision and are expected to be finalized in May, 1981. The outcomes are used for instructional purposes only: there are no statewide tests in any of the areas used for graduation requirements.

Contact:
Ms. Patricia Moran
Basic Skills Supervisor
Minnesota State Department of Education
550 Cedar
St. Paul, Minnesota 55101
(612) 297-2657

MISSISSIPPI:

No plans for assessing speaking and listening.

Contact:
Mr. Rex Pouncey
Supervisor, Accountability and Assessment Programs
State Department of Education
P. O. Box 771
Jackson, MS 39205

MISSOURI: In 1979, the State Department of Elementary and Secondary Education developed a "Guide for the Basic Essential Skills Test." The Guide describes three basic skills: reading/language arts, mathematics, and government/economics. Speaking and listening are not directly addressed, though there are four specific objectives under Reading/Language arts that relate to oral communication. These are: (number is objective number in Guide) 14) speak and write effectively in different social and business situations and with persons of varied ages or backgrounds, 17) recognize the main idea and specific details in an oral presentation, 19) communicate with persons of different age groups and in different situations by reading orally, and 21) follow oral directions to complete a process. The guide details sample criteria for these objectives, but the specific criteria used to judge the skill is the responsibility of the local school district.

Contact:
Dr. John J. Jones
Supervisor of Assessment
State Department of Elementary and Secondary Education
P. O. Box 480
Jefferson City, Missouri 65102

MONTANA: Nothing currently in speaking and listening.

Contact:
Claudette Johnson, Consultant
English/Language Arts
Office of Public Instruction
State Capitol
Helena, MT 59601

NEBRASKA: Currently state regulations call for assessment in reading, writing, and computation, but nothing in speaking and listening.

Contact:
Esther Arganbright
State Department of Education
Box 94987, 301 Centennial Mall South
Lincoln, NE 68509
(402) 471-2295

NEVADA: Current plans call for selection of instrument by 5/1/81 and yearly assessment beginning 10/81. Oral communication skills are not yet defined.

NEVADA: (continued)

Contact:
Herbert R. Steffens, Director
Planning and Evaluation
Nevada Department of Education
Capitol Complex
Carson City, NV 89710
(702) 885-5700

NEW HAMPSHIRE: New Hampshire completed an assessment of listening skills in October. Students in 5th, 9th, and 12th grades--about 600-700 students at each level--were assessed (multiple cassette messages, around 45 questions). No plans to do a random sampling assessment on speaking, but Dr. Baker (below) is working on 2-3 individual school district projects.

Contact:
Dr. Joanne Baker
Consultant, English Language Arts
State Department of Education
Division of Instruction
64 N. Main
Concord, NH 03301
(603) 271-3747

NEW JERSEY: Has plans to develop a speaking and listening assessment instrument sometime after this school year. No definite timetable as yet.

Contact:
Lydia Greenberg, Coordinator
State Testing Program
Department of Education
225 West State Street
Trenton, NJ 08625
(609) 792-5181

NEW MEXICO: Has a well developed plan for basic skills. However, basic skills are defined in the plan as reading, writing, arithmetic, and problem solving. No current plans to include speaking and listening.

Contact:
Howard J. Scheiber
Director of Writing Assessment and Pedagogy
Office of Basic Skills
Department of Education
Santa Fe, NM 87503

NEW YORK: Students are tested for final competency in grades 11 or 12. Final competency via the State Board of Regents exam is required for high school diploma. Speaking and listening are not identified as part of the exams which currently focus on reading, writing, and mathematics.

Contact:
Dr. Windsor Lott, Chief
Bureau of Elementary and Secondary Educational Testing
State Department of Education
Washington Avenue, Room 771
Albany, NY 12234
(518) 474-5099

NORTH CAROLINA: No plans to assess speaking and listening.

Contact:
Dr. William J. Brown
Director of Research
State Department of Public Instruction
Raleigh, NC 27611
(919) 733-3809

NORTH DAKOTA The state basic skills plan is still in process. The current draft does not specifically address speaking and listening, but focuses most of its content on reading and mathematics.

Contact:
Ethel Lowry
State Department of Public Instruction
State Capitol Building
Bismark, ND 58505
(701) 224-2292

OHIO: No statewide assessment or minimal competencies mandated. No one looking into speaking or listening.

Contact:
Jim Payton
Education Consultant, Planning and Evaluation
State Department of Education
65 South Front Street, Room 804
Columbus, OH 43215
(614) 466-4838

OKLAHOMA: No current plans to assess speaking and listening.

Contact:
Tommy Hodges
Administration, Office of Accreditation
State Department of Education
2500 North Lincoln, Room 245
Oklahoma City, OK 73105
(405) 521-3333

OREGON:

The Oregon State Board of Education adopted six instructionally-related state goals. The first goal states "Each individual will develop the basic skills of reading, writing, computation, spelling, listening, speaking, and problem solving." The Oregon Basic Skills Plan states it has very little data on oral communication. All school districts in Oregon are required to have written district, program and course goals in the areas of reading, writing, speaking, and computing. The local districts have quite a bit of flexibility in developing these goals.

Contact:
Dr. Gordon Asher
State Department of Education
700 Pringles Parkway, S. E.
Salem, OR 97310

PENNSYLVANIA:

One of the bright spots. The state has published a working copy of "Speech in the Classroom" which includes a well developed set of teacher rating scales for assessment of speaking skills. The assessment program is instructional and is designed to help the classroom teacher identify areas for instructional development in speaking skills. In addition, a companion document titled "Communication in the Classroom" suggests theoretical and curriculum materials for the classroom teacher. Both the rating scales and the curriculum materials are formed for different grade levels. The assessment procedures are primarily descriptive of types of communication instruction that might take place in the classroom. No plans to use this material for anything but instructional purposes.

Contact:
Evelyn W. Miller
Bureau of Curriculum Services
Pennsylvania State Department of Education
333 Market Street
Harrisburg, PA 17126
(717) 783-3947

RHODE ISLAND:

There are no plans for statewide assessment at the elementary level. At the secondary level, the area of language arts includes objectives for speaking and listening. A test-items bank includes sets of items (in the "life skills" area) that supposedly assess speaking/listening. Since these are very expensive to administer and score, there are no plans to use these items for a couple of years at least.

RHODE ISLAND: (continued)

Contact:
Martha Highsmith
Statewide Assessment Program
Department of Education
199 Promenade Street
Providence, RI 02908
(401) 277-3126

SOUTH CAROLINA: State has begun a process to implement an annual statewide assessment program for K-12. Kindergarten and primary grades currently have more oral skill items (both speaking and listening) than do later grades which focus more on reading and writing. Many of the items for the lower grades require an oral response. Elementary grades have been field tested, first full administration in Spring 1981. Secondary levels will be field tested in Spring 1981.

Contact:
Dr. Vana H. Meredith
Chief Supervisor
Educational Assessment Section
State Department of Education
1429 Senate Street, Room 604
Columbia, SC 29201

SOUTH DAKOTA: South Dakota has developed the "Language Arts Curriculum Guide: K-12," published in 1978. This guide is to be used by local districts and individual teachers to assist them in assessing and planning local language arts curriculum. The guide is divided into the following grade levels: Primary, K-3; Intermediate, 4-6; Junior High, 7-9; and Senior High, 10-12. Concerning listening and speaking, the guide states: "listening and speaking skills have been identified and emphasized for all the grade levels because they are the basis for developing the skills of reading and writing." Samples of Program Objectives for listening include: develops acceptable manners in listening (primary), recognizes similarities and differences in phonemes (primary), learns to listen for purposes of recall (intermediate), learns to value listening as an integral part of the communication process (junior high), etc. Examples of Program Objectives in speaking include: learns to participate without anxiety in a variety of verbal situations (intermediate), learns to use appropriate and varied vocabulary in oral presentations (intermediate), and develops fluency and poise in formal and informal speaking situations.

SOUTH DAKOTA: (continued)

Contact:
Kathy Hughes-Inan
Language Arts Program Director
Division of Elementary and Secondary Education
Kneif Building
Pierre, SD 57501
(605) 773-3508

TENNESSEE: No plans to assess speaking and listening.

Contact:
Dr. Tom Innes, Director
Educational Assessment
State Testing and Evaluation Center
University of Tennessee
1000 White Avenue
Knoxville, TN 37916
(615) 974-5385

TEXAS: Senate Bill 350 amended Sect. 16.176 of the Texas Education Code established a criterion referenced assessment program for reading, writing and math. It does not address speaking and listening.

Contact:
Mr. David Splitek
Division of Planning and Educational Assessment
Texas Education Agency
201 E. 11th Street
Austin, TX 78712
(512) 475-2066

UTAH: One of the few states that has a program of testing basic skills in place and operating begun in 1977 and adopted statewide, the system has a four by six matrix of skills. Four "life competencies" including consumer of goods and services, career, health and safety, and democratic governance. These four are applied to each of six basic skills including reading, writing, speaking, listening, computation, and problem solving. The student must demonstrate a competency in each of the 24 functional competencies defined in the matrix. A handbook provides material on curriculum locations, performance indicators, testing procedures and instructions, and required forms.

Contact:
Ferrin Van Wagoner, Coordinator
High School Programs
Utah State Board of Education
250 East Fifth South
Salt Lake City, UT 84111
(801) 533-5888

VERMONT: Another state with a program in place. The program began with the school year 1977-78. During this year 95,486 students were assessed to determine the level of mastery in the basic skills of reading, writing, listening, speaking, computing, and reasoning. The identified competencies will be required of graduates beginning with the class of 1981. The data and the discussion of results is contained in a series of booklets published by the Department of Education titled Basic Competency Program. Speaking and listening skills are evaluated by asking the student to orally respond to a variety of situations and then rated by their teacher.

Contact:
Shirley Reid
Director of Assessment/Testing
State Department of Education
Montpelier, VT 05602
(802) 828-3135

VIRGINIA: They are developing a statewide program called "Standards for Learning." This program includes learning objectives in all the basic skill areas, including speaking and listening. A tentative set of objectives were sent out for comment 12/80. Revised objectives should be published by summer of 1981. Programs will begin on a trial basis in the Fall, 1981. Assessment will probably take place beginning Fall 1982.

Contact:
Dr. Kenneth Magill, Supervisor
Jr. High/Middle Schools
Division of Humanities & Secondary Administration
Department of Education
P. O. Box 60
Richmond, VA 23216
(804) 225-2051

WASHINGTON: No current plans to test speaking and listening on a statewide basis. Basic skill program includes oral communication, but competencies are not defined.

Contact:
James Click, Coordinator
Basic Skills Program
Superintendent of Public Instruction
Old Capitol Building
Olympia, WA 90504

WEST VIRGINIA: No plans to assess speaking and listening.

Contact:
Mr. Charles D. Duffy, Coordinator
Assessment and Testing
State Department of Education
Est. Room B-057
1900 Washington Street
Charleston, WV 25305
(304) 358-3230

WISCONSIN: Wisconsin presently has a statewide assessment procedure for reading, writing, and mathematics. They are in the initial stages of developing a statewide assessment procedure in speaking and listening. The goals and specific skills to be assessed should be identified by 8/81, and a pilot test will take place approximately 1/82.

Contact:
Mr. Ben Brewer
Coordinator Basic Skills
Department of Public Instruction
125 South Webster Street
Madison, WI 53702
(608) 267-9268

WYOMING: Currently Wyoming has no plans for any statewide testing program in any area. Their Basic Skills Program, written in response to Sections 222 and 224 of Title II ESSA includes speaking, and listening but the plan does not specify competencies or evaluation.

Contact:
Alan Wheeler, Director
General Program Unit
Department of Education
Hathaway Building
Cheyenne, WY 82002

ASSESSMENT OF COLLEGE LEVEL SPEAKING AND LISTENING SKILLS

Rebecca B. Rubin
University of Wisconsin-Parkside

ABSTRACT

The aim of this paper is to describe the present conditions affecting assessment of speaking and listening skills at the college level. The paper covers the "state of the art" in college level programs, concentrating on the development of the Communication Competency Assessment Instrument. This instrument has undergone extensive reliability testing at the University of Wisconsin-Parkside. It is predicated on the Speech Communication Association's "Guidelines for Minimal Competencies in Speaking and Listening," which identifies four competency areas: communication codes, oral message evaluation, basic speech communication skills, and human relations. It also follows the SCA's "Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening Skills," making it one of few instruments that directly assess college level speaking and listening skills. Initial coefficient alphas support the reliability of the instrument; inter-rater reliability coefficients uphold the use of the rating booklet, which provides five skill levels for each of the 57 assessments (three assessments for each of 19 specific competencies). Short forms of the test are in their initial stages; judgments about the viability of using short forms are presented. In addition, conclusions about the need for college level assessment instruments are offered.

ASSESSMENT OF COLLEGE LEVEL SPEAKING AND LISTENING SKILLS

In March 1978, Assessing Functional Communication (Larson, Backlund, Redmond & Barbour, 1978) was published; this text detailed the state of the art in communication competency testing to date. The authors defined a number of main content areas, such as Developmental Language and Communication Skills, Communication Competence and Appropriateness, Receiving--Listening, Anxiety-Apprehension, and Interaction Descriptions. A total of 53 of the 90 tests reviewed were appropriate for college level or adult persons. However, no one test provided a comprehensive evaluation of speaking and listening skills; many of the instruments were specialized assessments of competencies such as anxiety, counseling/interviewing techniques, personal maturity, social competence, and group interaction.

At that time, the University of Wisconsin-Parkside had given up its search for an instrument to assess speaking skills, an area originally included in a comprehensive Collegiate Skills Program assessing writing, mathematics, reading, research paper development and library skills--all skills which students should have mastered by the time they have earned 45 college credits (Maeroff, 1978). The lack of an existing comprehensive assessment instrument in communication and concern over minority bias seemed to be the major reasons for "tabling" the speaking component. With a grant from the University of Wisconsin system (Rubin, 1980a), this author set out to discover the state of the art in college level competency testing, a prelude to developing a communication test for the UW-Parkside program. Few assessment instruments at the college level were found.

The American College Testing Program's College Outcome Measures Project (COMP) battery provided an assessment of general education knowledge and skills; one of the six areas assessed was labelled "Communicating," the "ability to send and receive information in a variety of modes (written, graphic, oral, numeric and symbolic), within a variety of settings (one-to-one, in small and large groups), and for a variety of purposes (for example, to inform, to understand, to persuade, and to analyze)" (Steele, 1979, p. 2). The purpose of the test was to assess effective functioning in adult roles in society. The "Speaking" portion of the Communicating area tended to concentrate on delivery and discourse; the test did include direct measures of speaking ability. However, the Collegiate Skills Program at the University of Wisconsin-Parkside was more concerned with the student's ability to function in college contexts, rather than in society at large.

Also examined was the program at Alverno College, which is more ambitious than most colleges can afford (Alverno College Faculty, 1976). At Alverno, communication ability is defined as the ability to write, read, speak, listen, use media, and use quantified data. The speaking portion assesses a student's ability to speak extemporaneously, clarify the setting and context, use effective delivery and linguistic conventions, organize and develop ideas, use visual aids, and evaluate his or her own strengths and weaknesses. There are 6 levels of student proficiency which can be attained in a combination of individualized and group testing sessions and courses. Again, a very ambitious program, but a broad mix of skill areas, some of which are not speaking and listening per se.

The Speech Communication Association's 1978 Task Force on Minimal Speaking/Listening Competencies presented the Minimal Competencies in Speaking and Listening for High School Graduates (Bassett, Whittington, & Staton-Spicer, 1978). This document seemed highly appropriate and relevant to this task of creating a college-level assessment instrument. It identifies four main competence areas: Communication Codes (ability to use and understand spoken English and nonverbal signs), Oral Message Evaluation (ability to use appraisal standards to judge oral messages and their effects), Basic Speech Communication Skills (ability to select and arrange message elements to produce spoken messages), and Human Relations (ability to maintain interpersonal relationships). These four main competence areas are further divided into 19 specific competencies, with examples of application within three contexts: occupational, citizenship, and maintenance. For purposes of the Communication Competency Assessment Instrument (CCAI), a fourth context--educational--was created, along with three application examples for each of the 19 competencies (see attached Table 1). These application examples were used as the basis for the first version of the CCAI. There were a total of 57 possible assessments to be made about a student's ability to function in specific educational environments: in class, and with instructors, fellow students, and academic advisors. (Table 1 - see appendix)

Meanwhile, the Speech Communication Association endorsed and published the Educational Policies Board's "Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening" (Backlund, Brown, Gurry, & Jandt, 1979); these criteria were followed in developing the CCAI. The Board document asserts that assessment instruments should be valid, reliable, and feasible. Specifically, it proposes the following guidelines: stimulus materials should require the demonstration of a skill; inferences about a speaking or listening skill should not be made from tests requiring reading and writing; the instrument should be unbiased; the test should assess skills occurring in familiar situations and in a variety of communication settings; tests should permit a range of acceptable responses; instruments should be standardized so that the test administrator's skills will not affect the results; the stress level should be equal to that of the setting in question; procedures should be practical in terms of cost and time and should involve simple equipment; and assessment should be suitable for the individual's developmental level.

At about the time of this publication, researchers from the University of Oklahoma were developing a communication assessment instrument to determine a student's admission into their business program (Scafe & Siler, 1979). They, too, followed the Speech Communication Association's lead on the competency areas, but opted for indirect assessment: that is, a written objective test. Direct assessment procedures were chosen as most appropriate for the CCAI.

Methods

The result was a 57-item version of the CCAI (Rubin, 1980b). There are three main sections to the test. The first asks the student to present a three-minute extemporaneous persuasive talk on a topic of

interest, during which the student's speaking ability is judged on many factors (e.g., volume, rate, clarity, gestures). Next, the student views a videotaped class lecture, and is then immediately questioned about the lecture and asked to respond in various ways to statements about his/her experiences in an educational environment. All student responses are either oral or nonverbal in nature; writing and reading abilities are not assessed. The test assesses only the student's ability to communicate through speech and nonverbal actions and to listen. At this stage of development, the CCAI took approximately one hour per student to administer.

This version of the test was critically examined by a number of persons and subjected to numerous refinements (Rubin, 1981). Content validity was achieved by presenting five communication faculty members with the 19 competencies and 57 assessment items arranged in random order; they were asked to place each assessment item into one of the 19 competency categories. Five questions that initially failed to meet the established 80 percent agreement standard were subsequently rewritten and correctly categorized by all evaluators. The questions were further refined, and a rating book with five levels of proficiency for each of the assessments was constructed, evaluated, and revised. Four faculty members trained as judges used the rating book to evaluate three students who had been videotaped earlier while taking the CCAI. A mean interrater correlation of .83 was attained, attesting to the reliability of the rating book. The communication skills of 77 students were assessed with this test version, each test session requiring about one hour.

As one might imagine, an hour's testing time per student presents monumental difficulties with large-enrollment institutions. Thus, the feasibility of reducing test size was investigated. Coefficient alpha analysis was performed on the 57-item CCAI. An overall alpha of .83 was achieved. To create a test of approximately one-third the number of assessment items (and one-half the administration time), the least consistent items were eliminated. That is, for each of the 19 competency areas, the most valid item was determined and the coefficient alpha for this 19-item short form was .79. It is extremely difficult to achieve a .80 coefficient alpha with less than 30 test items (Nunnally, 1978); thus, a .79 is considered respectable. However, to increase the coefficient alpha, four additional items were added to the 19-item short form, the four next highest ranking items, one in each of the four main sections of skills. The coefficient alpha for the 23-item form was .82. Additional testing was then necessary.

Eighty-three students who were enrolled in the basic communication course at the University of Wisconsin-Parkside completed the 23-item short form during February 1981 (see foctnote 1). A shortened version of the listening videotape stimulus was prepared. It had the same content of the original, minus unnecessary text (see footnote 2). All test administrations lasted 30 minutes or less. The data on these 83 students were combined with the data on the first group (77 students), and again submitted to coefficient alpha analysis. The coefficient alpha for the 23-item test was .80; for the 19-item test, the coefficient alpha was .78. It was determined that the two short forms were sufficiently similar not to warrant the inclusion of the four additional items which, in actuality, had no basis for inclusion. The following results pertain to the 19-item version of the CCAI.

Results

Table 2 shows frequency distribution of total scores on the 19-item test short form. The theoretical "passing" grade of 3.0 was used as a cutoff. The program proposed at UW-Parkside would declare all those receiving 57 total points (average grade of 3 for 19 items) as "competent"; those above 48 points (2.5 on a 5-point scale) as "in need of remedial self-paced work"; and those below 48 points "in need of formal training in a remedial course." The mean, median, and mode fell at about the same point, indicating a somewhat normal curve for the 160 students tested (see attached Table 2).

Initial analyses suggest that the instrument is free from bias. Male and female students did not appear to differ in their abilities to pass the test on the first trial ($X^2 = 0.68$, df = 1, n.s.). Similarly, no differences were found between minorities (hispanic, black) and non-minorities ($X^2 = 3.33$, df = 1, n.s.). However, there was a significant difference between minority ($\bar{X} = 56.13$) and non-minority ($\bar{X} = 63.74$) students ($t = 3.53$, df = 158, $p < .001$, two-tailed) on their total scores. At first glance, it appears as though the few minority students who did poorly on the test affected the mean of the minority student group (N = 16, s.d. = 9.75), more so than non-minority students who did not excel (N = 144, s.d. = 8.00). Additionally, the non-random sample of students reflects students who sometimes select the basic communication courses to improve known inadequacies in communication skills. This may have had an influence on the means.

Other data collected on the students, summarized in Table 3, included academic major, number of college credits completed, age, grade point average, number of communication or speech courses taken, and past speaking experience. Briefly, differences were found on the 19-item test for: (a) business and humanities majors; (b) those with 30 or fewer credits and those with over 60 credits; (c) the under-20 age group and the 20-24 age group; (d) those with a low GPA and those with a high GPA; and (e) those with little or no speaking experience and those with quite a bit of experience. Correlations of the raw scores of these data are found in Table 4. These analyses suggest that a variety of experiences and abilities leads to communication competence, as defined by the CCAI. Indepth investigation of these data is in progress (see attached Tables 3 and 4).

Additionally, each of the 19 items of the CCAI short form was closely examined. As Table 5 shows, 11 percent of the students tested had problems asking a question; 33 percent could not organize ideas well; 32 percent could not give accurate directions; 35 percent could not adequately express and defend a point of view; 10 percent didn't understand the difference between a fact and an opinion; 27 percent could not understand suggestions for improvement presented by an instructor; 14 percent could not adequately identify the work to be performed on an assignment when it was presented orally in class; and 49 percent could not describe the point of view of a person who disagreed with them (see attached Table 5).

Discussion

This initial research effort has disclosed that for some students, communication problems could very well inhibit learning. It is imperative that measures be developed to help students identify these problems and improve their communication skills so that they will not be disadvantaged learners. The State of Florida (Walker, 1979) is at work on this issue, as is Phil Backlund at Central Washington University. Valid and reliable procedures are needed to help students determine and correct their speaking and listening deficiencies.

Classrooms are communication arenas where students and teachers interact by communicating. Ability to survive in these environments is essential to successful college completion. The CCAI is proposed as a method of assessing the basic survival skills students need. The results of college-level competency testing can be used in a variety of ways.

It may be beneficial to some colleges to establish individualized instruction programs to help students improve specific skill areas. Other colleges may be able to identify specific courses which would help students improve their skills. And still other institutions could use the results in advising students about their weaknesses and helping them plan future course-work. Whatever the result, college-level competency assessment can provide useful information and feedback on students' skill achievement to help them become more effective communicators in their college classes.

FOOTNOTES

[1]This course is required only for Communication majors (Humanities); all other students choose the course as an elective. Extreme gratitude is expressed to Sally Henzl, graduate student at the University of Wisconsin-Milwaukee, for her assistance in this testing process.

[2]Professor Alan M. Rubin staged, directed and edited the videotape production and Professor Beecham Robinson participated in this production. Their assistance in this project is very much appreciated.

REFERENCES

Alverno College Faculty. Liberal learning at Alverno College. Milwaukee: Alverno Productions, 1976.

Backlund, P.M., Brown, K.L., Gurry, J., & Jandt, F.E. Criteria for evaluating instruments and procedures for assessing speaking and listening. In Resources for assessment in communication. Falls Church, Va.: Speech Communication Association, 1979.

Bassett, R.E., Whittington, N., & Staton-Spicer, A. The basics in speaking and listening for high school graduates: What should be assessed? Communication Education, 1978, 27, 293-303.

Larson, C., Backlund, P., Redmond, M., & Barbour, A. Assessing functional communication. Urbana, Ill.: ERIC, 1978.

Maeroff, G.I. Tests of minimum skills reaching college level. New York Times, December 26, 1978, pp. C1, C5.

Nunnally, J. Psychometric theory. New York: McGraw-Hill, 1978.

Rubin, R.B. Development of a communication competence assessment instrument for Parkside's basic skills program. In Enhancement of undergraduate teaching: Proposals and results. Madison: University of Wisconsin, 1980(a).

Rubin, R.B. The Communication Competency Assessment Instrument. Paper presented at the meeting of the Speech Communication Association, New York, November 1980(b).

Rubin, R.B. Assessing speaking and listening competence at the college level: The Communication Competency Assessment Instrument. Paper presented at the meeting of the International Communication Association, Minneapolis, 1981.

Scafe, M., & Siler, I.C. The development of a communication competency test at the University of Oklahoma. Paper presented at the meeting of the Central States Speech Association, St. Louis, 1979.

Steele, J.M. Assessing speaking and writing proficiency via samples of behavior. Paper presented at the meeting of the Central States Speech Association, St. Louis, April 1979.

Walker, N. (Ed.) Communication and computation skills necessary to the upper division: A report. Pensacola, Fla.: Pensacola Junior College, 1979.

Table 1

"Educational Context"

I. COMMUNICATION CODES

A. LISTEN EFFECTIVELY TO SPOKEN ENGLISH.
1. Understand directions given by a classroom instructor for class assignments.
2. Understand material presented in a class lecture.
3. Understand an instructor's suggestions for improving one's abilities.

B. USE WORDS, PRONUNCIATION AND GRAMMAR APPROPRIATE FOR THE SITUATION.

1. Use appropriate language in a classroom report.
2. Use appropriate grammar when speaking to others.
3. Use pronunciation which is understood by others.

C. USE NONVERBAL SIGNS APPROPRIATE FOR THE SITUATION.

1. Use appropriate gestures and eye contact when interacting with others.
2. Use appropriate facial expressions and tone of voice when conversing with one's instructor or fellow students.
3. Recognize and/or use appropriate gestures, eye contact, and facial expressions when communicating understanding or lack of understanding in a listening situation.

D. USE VOICE EFFECTIVELY.

1. Use appropriate rate when making a report in class.
2. Speak loudly enough to be heard in a classroom situation.
3. Use appropriate clarity when speaking with others.

II. ORAL MESSAGE EVALUATION

A. IDENTIFY MAIN IDEAS IN MESSAGES.

1. Identify the work to be performed when the assignment is given orally in class.
2. Recognize performance standards for work assigned orally in class.
3. Identify the main ideas in a class lecture.

B. DISTINGUISH FACTS FROM OPINIONS.

1. Recognize an opinion in a class lecture or report.
2. Recognize a fact in a class lecture or report.
3. Distinguish between facts and opinions in an interpersonal interaction.

C. DISTINGUISH BETWEEN INFORMATIVE AND PERSUASIVE MESSAGES.

1. Distinguish between informative and persuasive messages in a class report.
2. Distinguish between informative and persuasive messages in a class lecture.
3. Distinguish between informative and persuasive messages in an interpersonal interaction.

D. RECOGNIZE WHEN ANOTHER DOES NOT UNDERSTAND YOUR MESSAGE.

1. Recognize when an instructor or fellow classmate doesn't understand the question you are asking.
2. Recognize when an instructor or fellow classmate doesn't understand the question you are answering.
3. Recognize when others do not understand your explanation of a concept reported on in class.

III. BASIC SPEECH COMMUNICATION SKILLS

A. EXPRESS IDEAS CLEARLY AND CONCISELY.

1. Make a clear report on a subject of interest or one you've researched.
2. Concisely explain course requirements to a new student in class.
3. State clearly your reasons for taking a particular course.

B. EXPRESS AND DEFEND WITH EVIDENCE YOUR POINT OF VIEW.

1. Express and defend your view in a classroom report.
2. Express and defend your suggestions for improvements in your school.
3. Express and defend your position that a grade you received was incorrect.

C. ORGANIZE (ORDER) MESSAGES SO THAT OTHERS CAN UNDERSTAND THEM.

1. Use a chronological order to explain your activities throughout the day.
2. Use a topical order to explain a course you took last semester.
3. Use a problem-cause-solution order when discussing with an adviser/counselor an academic problem you are having.

D. ASK QUESTIONS TO OBTAIN INFORMATION

1. Obtain information about requirements for your major.
2. Obtain information about how to complete an assignment.
3. Obtain suggestions about how to improve your classroom performance.

E. ANSWER QUESTIONS EFFECTIVELY.

1. Answer an instructor's questions about your classroom performance.
2. Answer a question based on a class lecture.
3. Answer a question asked by a classmate about a course you are both taking.

F. GIVE CONCISE AND ACCURATE DIRECTIONS.

1. Direct fellow students in performing unfamiliar tasks or to an unfamiliar location.
2. Instruct a new student on how to do well in college classes.
3. Give accurate and concise directions to others.

G. SUMMARIZE MESSAGES.

1. Summarize oral instructions given by an instructor.
2. Summarize a class lecture.
3. Give a summary of students' suggestions to an instructor.

IV. HUMAN RELATIONS

A. DESCRIBE ANOTHER'S VIEWPOINT.

1. Describe the viewpoint of an instructor who disagrees with your evaluation of your classroom performance.
2. Describe the viewpoint of a fellow student who disagrees with your evaluation of a class you've both taken.
3. Describe the position taken on an issue by an instructor or fellow classmate with which you disagree.

B. DESCRIBE DIFFERENCES IN OPINION.

1. Describe differences in opinion with fellow students about course related issues.
2. Describe differences in opinion about the steps necessary to accomplish your academic or vocational goals.
3. Describe differences of opinion which occurred in a class discussion.

C. EXPRESS FEELINGS TO OTHERS.

1. Express satisfaction or dissatisfaction to an instructor about a course you have taken.
2. Express feelings of satisfaction or dissatisfaction about working with others on group projects for classes.
3. Express empathy to a friend who has not done well on a class assignment or in a course.

D. PERFORM SOCIAL RITUALS.

1. Introduce yourself at the beginning of the semester in class.
2. Request an appointment with a counselor or adviser.
3. Conclude a conversation with an instructor.

TABLE 2

RELATIVE FREQUENCY OF 160 STUDENTS ON THE 19-ITEM SHORT FORM

	Score	Frequency	Standard Deviation	Score on 5-pt. scale
	39	1		
	40	1		
N = 7	44	1		
	45	1		
4.375%	46	1	-2	
	47	2		2.5
	48	1		
	49	3		
N = 33	50	1		
	51	3		
20.625%	52	1		
	53	5		
	54	5		
	55	4	-1	
	56	10		3.0
	57	3		
	58	6		
	59	3		
	60	3		
	61	9		
	62	3		
*	63	8	$\bar{X}$	
**				
***	64	18		
	65	7		
	66	4		3.5
N = 120	67	8		
	68	6		
75.000%	69	1		
	70	13		
	71	4	+1	
	72	5		
	73	4		
	74	4		
	75	2		
	76	1		4.0
	78	3		
	79	2		
	80	1	+2	
	82	1		
	83	1		

*Mean = 62.98 **Median = 63.78 ***Mode = 64.00

s.d. = 8.475 range = 44.00

TABLE 3

ANALYSIS OF VARIANCE TABLES FOR THE 19-ITEM TOTAL SCORE

	Mean*	N		df	MS	F	p
ACADEMIC MAJOR							
Science	65.50	10	between				
Humanities	65.95_a	39	groups	8	170.31	2.56	.012
Fine Arts	60.80	5					
Behavioral Sci.	65.60	10	within				
Business	59.75_a	44	groups	151	66.61		
Computer/Eng.	67.60	5					
Social Science	69.00	5					
Education	63.00	3	total	159.			
Undecided	61.26	39					
COLLEGE CREDITS COMPLETED							
0-15 credits	61.12_a	86	between groups	4	390.26	6.14	.0001
16-30 credits	61.30_b	23					
31-45 credits	65.00	13	within groups	155	63.61		
46-60 credits	63.00	12					
Over 60 credits	69.62_{ab}	26	total	159			
AGE							
Under 20	60.88_a	92	between groups	4	252.50	3.76	.006
20-24	65.53_a	36					
25-29	64.73	11	within groups	155	67.17		
30-39	67.54	13					
Over 39	65.88	8	total	159			
GRADE POINT AVERAGE (4-pt. scale)							
Under 2.00	59.42_a	24	between groups	4	223.10	3.72	.007
2.00--2.49	63.39	23					
2.50--2.99	64.12	42	within groups	122	59.96		
3.00--3.49	67.50_a	28					
3.50--4.00	65.90	10	total	126			
COMMUNICATION COURSES							
None	61.01	71	between groups	5	841.89	2.45	.036
1	63.12	49					
2	64.58	19	within groups	154	68.69		
3	67.70	10					
4	66.83	6	total	159			
5 or more	69.40	5					
PAST SPEAKING EXPERIENCE							
None	61.06_{ab}	49	between				
Classes other than speech	60.35_c	26	groups	5	326.68	5.14	.0002
Speech Classes	62.28_d	46	within				
Some Experience outside of speech class	66.52_a	25	groups	154	63.56		
Good amount outside class	72.50_{bcd}	10	total	159			
A lot outside class	65.75	4					

*Means with a common letter in the subscript differ at or beyond the .05 level of confidence utilizing the Tukey-B procedure.

TABLE 4

CORRELATIONS OF DEMOGRAPHIC VARIABLES AND THE 19-ITEM SHORT FORM

	CC	Age	GPA	CCC	SE	Race	Sex
Credits Completed							
Age	.34***						
Grade Point Average[1]	.31***	.29**					
Communication Courses Completed	.34***	.19*	.11				
Speaking Experience	.21**	.15	.06	.58***			
Race[2]	-.04	-.02	-.32***	-.06	-.02		
Sex[3]	-.08	.18*	.02	.06	.07	-.01	
19-Item Short Form	.35***	.21**	.28**	.28***	.31***	-.27***	-.02

*p < .05 **p < .01 ***p < .001

[1] 33 students had no GPA accumulated; they were labelled "missing" for this analysis. N = 127 for all GPA correlations. For all other correlations, N = 160.

[2] For the Race variable, 0 = Non-minority, 1 = Minority.

[3] For the Sex variable, 0 = Male, 1 = Female.

TABLE 5

PERCENTAGE OF STUDENTS' SCORES ON A 5-POINT SCALE
FOR THE 19-ITEM SHORT FORM

	lowest				highest
Item	1	2	3	4	5
Pronunciation	1.2%	0.0%	14.4%	47.5%	36.9%
Facial Expr./Tone of Voice	0.0	9.4	29.4	41.2	20.0
Speech Clarity	1.2	15.0	36.2	43.1	4.4
Persuasive/Inform. Distinct.	5.0	20.6	30.6	31.9	11.9
Clarity of Ideas	1.9	12.5	36.9	39.4	9.4
Express & Defend Pt. of View	7.5	27.5	38.7	18.1	8.1
Recognize Non-Understanding	3.1	2.5	23.7	46.2	24.4
Fact/Opinion Distinction	4.4	5.6	23.7	61.2	5.0
Listening (Understand Suggestions)	5.0	22.5	47.5	20.6	4.4
Identify Main Ideas	3.7	10.0	38.1	28.7	19.4
Summarize	8.1	7.5	36.9	22.5	25.0
Social Ritual	1.9	10.6	30.0	40.6	16.9
Ask Questions	1.9	9.4	37.5	34.4	16.9
Answer Questions	1.2	8.7	55.6	26.9	7.5
Express Feelings	5.6	29.4	19.4	31.9	13.7
Organize Ideas	20.6	12.5	39.4	26.2	1.2
Give Directions	12.5	20.0	30.0	20.6	16.9
Describe Another's View	17.5	31.3	28.7	18.8	3.7
Describe Diff. in Opinion	14.4	30.0	34.4	16.9	4.4

POTENTIAL SOURCES OF BIAS IN SPEAKING AND LISTENING ASSESSMENT

Richard J. Stiggins
Clearinghouse for Applied Performance Testing

ABSTRACT

There is some evidence to suggest that the use of performance assessment can, in certain educational assessment contexts, minimize the adverse impact of testing on minorities. Performance assessment, an alternative to traditional paper and pencil testing, relies on the direct observation and rating of behavior by qualified judges. For obvious reasons, such assessment plays a key role in the measurement of speaking and listening skills. Because patterns of oral communication are greatly affected by potentially divergent linguistic and cultural factors, great care must be taken to deal with potential sources of bias in the assessment if valid and reliable results are to be achieved. This paper reviews those potential sources of bias, and suggests strategies for dealing with each. Purposes and strategies for assessing speaking and listening skills are reviewed, as are the characteristics of good speaking and listening tests. Specific sources of bias in speaking and listening assessment are identified in (a) test administration environment, (b) stimulus conditions, (c) response modes, and (d) rating or scoring procedures. Specific suggestions are offered for planning and conducting a speaking and listening assessment that precludes problems of bias.

POTENTIAL SOURCES OF BIAS IN SPEAKING AND LISTENING ASSESSMENT

A test is regarded as biased when some characteristic of the test interacts with some characteristic of the test taker in such a way as to distort the meaning of the results for a certain group of examinees (Shepard, 1980). Such distortions can lead to invalid assessments and inappropriate educational decisions. For example, test items use language or syntax more readily comprehensible to members of one group—with its cultural and linguistic experience—than to another group. Or the exercises may be based on problem situations with which one group has simply had more familiarity than another. In both cases, examinees from each group might be equally capable of performing competently on the test, but some language or experience factor extraneous to the trait or skill being measured might distort test results and thus lead to an invalid assessment.

This paper explores some of the potential sources of bias in speaking and listening assessment. Because speaking skills and listening skills call for different measurement strategies, they are treated separately. Each is described in terms of assessment purpose and strategies, characteristics of a good test, potential sources of bias, and strategies for avoiding bias.

Assessing Speaking Skills

Purpose and Strategies. The purpose here is to measure the examinee's ability to communicate orally. Assessors have two measurement options: (1) a paper and pencil objective test measuring knowledge of the principles of good communication, or (2) direct observation of examinees' ability to apply those principles in a real speaking context.

Although knowledge of effective communication principles may correlate highly with actual speaking proficiency, most would contend that direct observation of an examinee's performance yields more dependable conclusions regarding speaking proficiency. In short, the more effective strategy is to elicit a sample of speech in a real or simulated situation, and to ask a trained rater to evaluate results according to prespecified performance criteria. Speaking proficiency can be rated holistically or it can be subdivided into such components as organization, content, and expressiveness, and rated analytically—depending on the assessment purpose or the information desired.

Characteristics of a Good Test. Test characteristics necessarily vary somewhat as a function of testing purpose. All effective speaking tests, however, share some common characteristics: (1) clearly focused, carefully prepared speaking exercises; (2) multiple realistic (real or simulated) speaking contexts; (3) clearly specified (in writing) performance criteria; (4) effective rater training; and (5) a carefully conducted rating process.

Testing purpose determines the level of information required in test results. For example, placement into remedial speech courses typically calls for information on overall communication proficiency. On the other

hand, diagnosis of strengths and weaknesses for purposes of designing individualized instruction requires thorough analysis and scoring of the component skills of oral communication.

Potential Sources of Bias. There are at least four potential sources of bias in a performance test of speaking skill: the test administration context, the stimulus situation, the response mode, and the scoring process.

Bias can come from the test administration environment, if that environment is more familiar or comfortable to one cultural group than another. Environmental factors that can affect performance include amount of anxiety with the test format, satisfaction with physical surroundings, frequency and nature of distractions, and race or sex of the test administrator. When these factors affect one group's scores more than another's, they can lead to invalid conclusions regarding speaking skill.

The exercises (or stimulus) exercises can also be biased if based on inappropriately narrow cultural perspectives. Competent examinees who lack understanding of or experience with item content may be inappropriately judged incompetent.

Response bias can arise as a result of cultural differences in the nature and form of oral communication typically used by examinees. For example, members of certain cultural subgroups may be uncomfortable with formal speaking contexts. Though competent speakers, such persons may experience sufficient anxiety or lack of motivation in an unfamiliar situation, or one perceived by their culture as inappropriate, to distort test results.

And finally, bias can arise in scoring whenever raters are called on to make subjective judgments. The potential for rater bias is a function of the rater's social experience and attitudes, the clarity and precision of the scoring criteria and standards, and the extent to which the rater has internalized those standards. Guilford (1954) has outlined the many forms of bias in ratings, and others (Schmitt and Lappin, 1980; Campbell, 1972; Quinn, 1969; and Brignoli, et al., 1979) have shown how these factors influence performance ratings of different cultural subgroups. Rater bias is perhaps the most significant threat to test score dependability in performance assessment.

Avoiding Bias. The general strategy for avoiding bias in speaking assessment is to eliminate those factors that can distort the meaning of test scores. For example, the assessment can be administered in a familiar and comfortable environment by a test administrator who is not likely to be intimidating or otherwise distracting to examinees. Stimulus exercises can be reviewed in advance by members of appropriate cultural subgroups to ensure relevance and familiarity to all examinees. The use of formal and informal speaking exercises which are within the real-world experience of all examinees can reduce response bias and anxiety, and increase motivation to perform. Another strategy for controlling anxiety is unobtrusively assessing speaking skills via direct observation of examinees in everyday communication settings. And finally, rater bias can be controlled through the very careful specification and review of performance criteria, the careful selection of raters and thorough rater training.

Assessing Listening Skills

Purpose and Strategies. The purpose here is to measure an examinee's ability to comprehend oral communication. As with speaking assessment, the examiner has two options: a test that measures knowledge of the principles of good listening, or a performance test that measures ability to apply those principles.

The most effective strategy is typically to use a performance test that relies in part on objectively scored test items of listening comprehension. That is, the listening exercise is presented orally in a real or simulated context, and the examinee's comprehension of the message is tested using traditional multiple choice items. As an alternative to written items, one can ask the examinee to paraphrase all or part of the message, and rate the level of comprehension conveyed in that paraphrase according to prespecified criteria.

Characteristics of a Good Test. A good listening test is one that provides the kind of information needed for appropriate educational decisions. Mead (1978) has pointed out that such a test focuses on skills central to the listening process by using real verbal communication presented orally, rather than using reading test passages. Mead urges the use of short, interesting stimulus materials that are relevant to the communication context and experience of the examinees. In addition, Mead contends that a good test should not confound listening and reading skill by requiring that examinees read test items after listening to passages. Instead, test items should be presented in both an oral and written form. And finally, Mead suggests the use of items that require (1) recalling significant details, (2) comprehending main ideas, (3) drawing inferences, (4) making judgments about speakers, and (5) making judgments concerning information presented.

Raters who are asked to judge listening proficiency based on an examinee's paraphrasing skill must be carefully trained and must have clearly specified performance criteria on which to base their judgment.

Potential Sources of Bias. As with speaking assessment, sources of bias in listening assessment can include the test administration environment, stimulus, response mode and scoring procedures. If the test administration situation is unfamiliar or uncomfortable to certain cultural groups (more than others) or the stimulus materials are based on narrow cultural or linguistic experiences or perspectives not common to all examinees, the test can yield biased results.

The response mode can affect bias, too. If objective test items are used to measure comprehension, such factors as differences in test wiseness, anxiety or motivation to do well can lead to biased judgments. If paraphrasing is used to measure comprehension, speaking and listening proficiency may be confounded. Cultural differences in speaking style may also lead to biased assessment. Whenever factors extraneous to the actual listening ability of examinees influence test scores, and do so to differing extents across cultural subgroups, the meaning of test scores will be distorted for members of some groups and the test scores will lead to biased judgments about listening proficiency.

And finally, scoring procedures can contribute to test bias when subjective ratings of paraphrased messages are the basics for determining proficiency. Rater bias is, of course, not a problem with objective test items.

Avoiding Bias. The key to avoiding bias in listening assessment is to use familiar, realistic and relevant samples of oral communication combined with high quality assessment procedures. Careful review of stimulus materials, items and rating criteria by members of appropriate cultural subgroups prior to test administration can increase quality and help eliminate bias. Test items should be prepared by qualified testing specialists, and subjected to careful pretesting and psychometric analysis prior to actual administration. Bias in the response mode because of differences in anxiety and motivation can be reduced by using paraphrasing as an alternative to the traditional objective test. These gains, however, can be offset by rater bias unless raters are thoroughly trained and given carefully specified criteria on which to base their judgments.

Conclusion

In speaking and listening assessment, test environment, stimulus conditions, response mode, and scoring procedures all influence the meaning and utility of test results. If performance tests are carefully developed by persons knowledgable about and sensitive to the cultural and social experiences and traditions of the groups to be tested, such tests can provide useful, meaningful scores. The use of stimulus circumstances and exercises equally familiar to different groups can help ensure that everyone has a chance to perform successfully. When members of different groups are faced with a test stimulus or response mode that prevents them from demonstrating their ability, the test opportunity is biased. The use of performance rating procedures not based on explicit criteria, and not applied by carefully trained judges, can lead to rater bias. Considerable care must be taken to prevent both opportunity and rater bias, since either can significantly influence the dependability of speaking and listening assessment results.

REFERENCES

Brugnoli, G.A., J.E. Campion and J.A. Basen. Racial bias in the use of work samples for personnel selection. Journal of Applied Psychology, 1979, 64(2), 119-123.

Campbell, J.T. An investigation of sources of bias in the prediction of job performance: A six-year study. Princeton, NJ: Educational Testing Service, 1972.

Guilford, J.P. Psychometric methods (2nd ed.), New York, NY: McGraw-Hill, 1954.

Mead, Nancy A. Issues related to assessing listening ability. Denver CO: National Assessment of Educational Progress, 1978. 16 pp.

Quinn, J.L. Bias in performance appraisal. Personnel Administrator, 1969, 32, 40-43.

Schmitt, N. and M. Lappin. Race and sex as determinants of the mean and variance of performance ratings. Journal of Applied Psychology, 1980, 65(4), 428-435.

Shepard, L.A. Definitions of bias. A paper presented at the Johns Hopkins University National Symposium on Educational Research, Washington, D.C., November 7, 1980.

USING PERFORMANCE RATING SCALES IN LARGE SCALE ASSESSMENTS OF ORAL COMMUNICATION PROFICIENCY

Donald L. Rubin
The University of Georgia

ABSTRACT

Speech performance rating procedures optimize both the informativeness and the pedagogical "washback" of communication competency assessments. Rating procedures can be adapted to a variety of communication situations, and may serve a number of administrative purposes. Appropriate selection of rating instruments and criteria is determined by the performance situation and the purpose of the test. Considerations of content validity bear on choices of performance situations as well as rating criteria. Considerations of reliability bear on the size, number, and variety of performances sampled as well as on agreement between raters. Considerations of feasibility bear on the time and personnel needed to elicit and rate speech performances.

USING PERFORMANCE RATING SCALES IN LARGE SCALE ASSESSMENTS OF ORAL COMMUNICATION PROFICIENCY

Federal legislation presently includes oral communication in its enumeration of basic skills (ESEA Part 162, Federal Register, April, 1979, 25151). Many state and local education agencies have already begun to test oral communication competencies, while others have initiated preliminary planning in this direction (Backlund, 1981). This paper does not argue for large scale testing of oral communication proficiency. Indeed, it is preferable not to test when an assessment program threatens to undermine instruction. This paper is, however, premised on the fact that programs for testing oral communication competencies are upon us. It is therefore critical to examine available assessment methods in order to select those which will maximize the benefits of such programs.

Great Britain, where large scale testing of oral skills has been conducted for some time, provides a valuable example of the impact of oral assessment on education. British Certificate of Secondary Education examinations have included some type of speech sample since 1965 (Hitchman, 1966; Schools Council, 1966). Although curricular documents have stressed the importance of "oracy" (Bullock, 1975), British educators recognize that the institution of formalized testing lent respectability to speech instruction (Wade, 1978) and shaped its nature (Wilkinson, 1968). Barnes (1980, p. 125) observes that "any monitoring of oracy during secondary schooling will be proposing a wider range of curricular concerns in oracy than schools presently undertake...Thus, in secondary schools at least, the monitoring of oracy is likely to be leading practice in schools rather than responding to it."

With an eye toward effects on pedagogical as well as psychometric adequacy, then, this paper examines one type of communication assessment procedure: speech performance rating. It constructs a rationale for this procedure and describes conditions influencing selection of particular speech performance rating techniques. The paper also addresses the manner in which considerations of content validity, reliability, and feasibility can affect the adequacy of speech performance rating as it is likely to be implemented in school settings.

Rationale for Speech Performance Rating

Speech performance rating encompasses assessment procedures in which student speech is sampled in some more or less communicative context. Judges rate the speech samples according to some criteria of communication quality. Speech performance rating thus excludes more narrowly directed testing of linguistic ability: for example, tests of articulation requiring students to read word lists, or oral tests of grammar which employ criteria making reference only to syntactic patterns. More to the point of this essay, speech performance testing contrasts with indirect tests of speaking proficiency wherein students are questioned about oral communication but do not engage in it.

Indirect tests are advantageous because they are easily administered and easily scored. Though indirect tests of skill surely lack face validity, they may indeed be highly valid indicators or predictors of competence to perform. As an example, consider multiple-choice tests of writing ability (Breland, 1977; Godshalk, Swineford & Coffman, 1966; Palmer, 1961). Speaking competence, however, has proved resistant to indirect testing. One notable effort met with difficulty in constructing adequate items (Mead, 1977). Communicative intentions are complex and communicative contexts are multidimensional. Meaning can sometimes be conveyed effectively by deviating from norms. For example, a speaker can, under certain circumstances, express endearment by uttering, "You're such a jerk, Jackson." Competent speakers can envision conditions under which nearly any supplied alternative could appropriately serve some communicative intention, and it is no mean task to sufficiently circumscribe a communicative situation within the confines of an examination question. Indeed, Howie-Day (1977) found no age-dependent patterning in choice of supplied persuasive appeals, but only in rationales used to explain those choices.

Indirect tests of speaking ability represent a challenge to test construction, but more importantly, they also represent a threat to acceptable instruction in oral communication. Wilkinson (1968) described the manner in which the mode of oral assessment has a "washback" effect on the manner in which oracy is taught. Similarly, Rubin (1980) discussed the "pedagogical validity" of oral assessment procedures which is ascertained by determining what changes in (a) curricular content and (b) instructional practices would maximize test scores. If indirect testing of speaking ability were instituted for decisions of major consequence, it is likely that students would accordingly soon be learning <u>about</u> communication instead of learning <u>how to</u> communicate. The experiential component which constitutes the core of most communication training would atrophy. The history of Educational Testing Service's College Board English Composition Test provides an instructive parallel. Test constructors developed a reliable and cost-effective instrument which predicted writing sample ratings with great accuracy. Original writing samples were consequently eliminated from the College Board Composition Test. However, English teachers protested that the indirect test caused a reduction in writing time-on-task and an increase in punctuation and usage drill. Ultimately this pressure resulted in reinstitution of some type of writing sample (Educational Testing Service, 1979).

The limits of test construction and the demands of acceptable instruction, then, point to the need for direct performance testing of speaking skill. Still, it may be argued, rating procedures may not be the most efficient means of assessing quality from speech samples. The most viable alternative to rating is some form of discourse feature analysis. That is, speaking skill may be inferred on the basis of specific structures, strategies, or quantifiable items appearing in speech samples.

Both Loban (1976) and McCaleb (1979) suggest syntactic complexity as a discourse feature from which evaluations of communication competence can be drawn. But while syntactic complexity is generally an index of syntactic maturity, it is also affected by a variety of contextual factors, and in no way directly reflects quality of expression

(Crowhurst, 1979). Indeed, in writing examinations, the most able writers may produce the least complex syntax (Rosen, 1969). In a like vein, students who participate in instruction in referential communication accuracy often produce briefer, denser messages (Fry, 1969; Shantz & Wilson, 1972). But no one would consider such telegraphic speech to be qualitatively superior communication. Context-creating statements were taken as a reflection of audience adaptedness in one study (Bracewell, Scardamalia & Bereiter, 1978), but were shown unrelated to age or expertise in another (Rubin & Piche, 1979). In short, speech is a wondrously varied and flexible tool, and any equation absolutely linking specific discourse features with quality of expression is doomed to disconfirmation.

Speech performance rating—cumbersome, liable to measurement error, relatively costly—appears to be the optimal method for assessing speaking proficiency. It has no inherent impediments to validity; it is clearly relevant to the target skill domain. Moreover, speech performance rating may have a salubrious effect on instruction. Though challenging, it is feasible; its technology is currently available.

Situations for Performance Rating

An emphasis on speech situations can be misconstrued as an emphasis on speech types: the persuasive speech, the panel discussion, the dramatic reading. This interpretation is misleading because it oversimplifies the dynamics of communication behavior and can lead to mechanical, rote preparation and performance. A more adequate framework refers to communication functions realized in the context of particular audiences, particular tasks or topics, and particular settings. We communicate to inform, to express our emotional state, to regulate others, to persuade, to bond with others, and so on (Allen & Brown, 1976; Jacobson, 1960). (And interactions generally serve several of these functions simultaneously). Moreover, we execute these functions differently under varying circumstances. A formal presentation about competency testing, for example, will be adapted in a number of respects if it is being delivered to a school faculty as compared to a civic group, and adapted again differently if delivered from behind a podium as compared to from a seat at a round table.

Accordingly, evaluation criteria need to reflect the peculiar demands of speech situations. It is not reasonable to judge, say, an interview performance by means of the same criteria as are applicable to a peer discussion. Similarly, an informative speech calls for standards different from those for a ritual speech of appreciation. By the same token, genteel pronunciation and usage typical of dominant cultural groups may be appropriate in some situations, but not in others.

Rating procedures can account for situationally appropriate adaptation in several ways. Some scales include an item explicitly citing appropriateness or adjustment to audience and purpose. In another approach, the rubrics describing levels of quality will be different for different speech situations. For example, a high rating along a dimension of "intelligibility" may require avoidance of technical jargon when addressing a heterogeneous audience, but allow specialized vocabulary when speech is directed to a specialized audience. Finally,

different speech situations may require the application of altogether different criteria, reflecting divergent rhetorical demands of those situations (Lloyd-Jones, 1977). Thus, an item "shows initiative" may be included in a scale for rating simulated job interviews, while an item "refutes counter-arguments" would be fitting only in evaluating a persuasive performance.

Speech situations can be arrayed along a continuum from formal, extended, uninterrupted, and planned on the one hand to informal, reciprocal, spontaneous and nondirected on the other (Wilkinson, 1968). More formal speech situations yield greater control over the examination and greater ease of rating. These situations may not be of the highest life-role utility, however. In addition, formal speech situations may be foreign and threatening to large numbers of students, resulting in depressed levels of performance. These situations may be especially alien to members of cultural minorities (Gay & Abrahams, 1973; Philips, 1970).

Because performance appropriately varies as a function of situation, and because different situations may tap divergent constellations of skills, an ideal assessment program in this domain would sample performance in a variety of situations. This paper returns to the problem of sampling speech situations in a later section on content validity, and addresses the question of equivalence of performance across situations in a section on reliability. Some administrative options providing for a variety of speech situations in an assessment program may be briefly mentioned at this juncture, however.

Utilizing a single speech situation obviously simplifies administration and reduces costs. In one program, initial testing takes place in the context of a persuasive public address, while retesting of those who fail to achieve criterion performance levels takes place in the context of a simulated job interview (Rubin & Bazzle, in preparation). This arrangement requires that the reliability and difficulty of equivalent forms be verified. Alternatively, individual students may be observed only in a single speech situation (self-selected or randomly assigned), with that situation drawn from a list of testing options (e.g., Gary Community Schools Corporation, 1977-1978). Again, equivalency among these options must be established. Finally, students may be subjected to multiple observations representing a range of situations. This procedure is administratively feasible only if it is conducted by regular classroom teachers and incorporated into classroom instruction, at least for initial screening (e.g., Massachusetts Department of Education, 1979; Vermont Department of Education, 1978). Use of inadequately trained classroom teachers as raters, however, risks considerable unreliability.

Reading Aloud. This most formal of speech situations affords great ease of administration and scoring. For this reason it is the most common situation employed in British Certificate of Secondary Education examinations (Barnes, 1980; Hitchman, 1966, 1969; Wilkinson, 1968). It should be noted that tests composed of reading aloud can reflect more

than ability in vocal delivery—articulation, pronunciation, and the like. Hitchman (1966) proposes that prose and poetry reading be judged by the criterion of "interpretation," which signifies skill at conveying to a listener the mood and meaning of a passage.

Extended Monologue. Everyone has a notion of what it means to give a speech. Unfortunately, the conditions for speech making—an authentic purpose and a potentially receptive audience—are frequently absent in classroom practice, much less in oral competency testing. Hence the term "extended monologue." Extended monologue assessments are most adequate when they make provision for at least a simulated audience and purpose. In one program, for example, students conduct a simulated public hearing before a Board of Education. Each student makes a presentation before the Board, taking a position on one of three agenda items (Rubin & Bazzle, in preparation). Students seem to perform better when they are dealing with personally meaningful subject matter in a well-defined situation. Instructions that read "Speak for three minutes on the proper role of American foreign aid" are not likely to elicit maximal performances from most youngsters.

Group Discussion. Group discussion formats span a range of speech situations from panel presentations—which may be highly formal—through free discussion. Group discussion tasks deserve consideration in oral assessment programs first, because peer interaction may optimize some students' performance; second, because small-group problem solving has value as an instructional technique across the curriculum; and third, because small-group communication skills are of high utility in life-role contexts. Group discussion assessment tasks entail some especially thorny challenges to measurement, however (Keltner, 1953). The quality of individual discussants' performance is affected by the quality of other members' performances. Different group roles—leader, for instance—require the application of different evaluation criteria. When raters take active roles in directing discussion, they can alter the quality of students' performances in ways which are extraneous to students' communication skills (Barnes, 1980). Still, it appears that discussion skills can be rated in a reliable manner (Becker, 1956). Criteria mentioned in conjunction with assessment of group communication performance emphasize cooperative thinking, integration into the group, and appropriate degree of participation (neither reticent nor dominating) (Becker, 1956; Crowell, 1953).

Interview. Bazen (1978) rightly observes that under conditions of testing, dyadic communication never qualifies as true conversation. When a rater interacts with an individual testee, the situation is structured and goal oriented; it is an interview. Some assessment procedures attempt to circumvent this state of affairs by simulating conversational performance through role-playing (Massachusetts Department of Education, 1979). Others engage students in direct interviews about autobiographical information (Gary Community School Corporation, 1977-1978). Simulated job interviews have also been used for oral assessment (Rubin & Bazzle, in preparation). Tests of second language proficiency (Mullen, 1978) concur with experience in testing communication proficiency (Bazen, 1978; Hitchman, 1966) in suggesting that the interviewer/rater can seriously affect students' performance scores.

Purposes for Speech Performance Rating

Information derived from large scale performance assessment may serve as input for several types of educational decisions. Test results may be used to certify the competency of individual students for purposes of promotion or graduation. The results may be used to place students in specific instructional sequences in order to remediate deficiencies or enrich existing strengths. Or, assessment information can be used to evaluate the effectiveness of programs or of teachers.

The intended purpose of a speech proficiency test will determine, in part, the type of rating scale needed. Instruments designed for the purpose of certifying communication competency must somehow reflect those oral competencies previously identified by educational agencies. This can be accomplished most satisfactorily by constructing scale items corresponding to each competency. For complex competencies which have not been defined with precision (e.g., "Deliver a simple speech to inform about a personal interest or hobby") it may be necessary to construct scale items corresponding to subskills underlying the more general objectives ("uses introduction," "formulates a logical sequence of ideas," "oral delivery is intelligible," etc.). If the resultant rating scale is deemed too unwieldy, it is possible to create a simpler, more impressionistic scale composed of fewer items. For such simplified scales, rubrics describing indicators for each level of quality will express the same competencies or subskills as would appear on a more complex instrument. The simplified instrument may seem ostensibly uninformative; students must be given an explicit list of rating criteria in order to adequately prepare and to interpret their results.

When test results are to be used for the purpose of student diagnosis and placement, there is little alternative to a highly detailed rating instrument. Scale items will explicitly state evaluative criteria corresponding to instructional goals. Thus, a student who rates poorly on "ideas sufficiently elaborated" may participate in an instructional module training invention skills through practice in use of heuristics, brainstorming, and use of forms of support. That same student may not require the additional training in organization skills or in kinesic delivery skills which may be indicated for another student.

The effectiveness of a program or of a teacher is generally ascertained by assessing the product of that program or teacher: student achievement (though this approach is useful only when student aptitude and institutional resources are also taken into account). Rating scales for use in summative evaluation of achievement need not explicitly state detailed criteria if the ultimate users of the test results are simply interested in documenting instructional impact. More detailed information will be required if test users are interested in "fine tuning" existing programs. Of course raters will need to be well versed in the specific evaluation criteria, in any event, if they are to judge students' oral performances in a consistent fashion.

Types of Speech Performance Rating Scales

It is not possible to recommend a single type of rating scale for use in speech proficiency assessment. Instruments vary in the degree to which items represent detailed competencies or subskills. They differ

also in the degree to which criteria are tailored to particular demands of communication tasks. In order to select an appropriate rating scale, testing agencies must consider the purpose of the assessment program and the speech situations in which students' oral performance are sampled.

Type-of-Speech Checklist. Some speech assessment programs seek to incorporate evaluation into normal classroom activity. This approach may mitigate student anxiety and also relieve (or shift) some administrative burdens. Also, it allows for testing in a variety of speech situations. The major disadvantage of such schemes is lack of control over administration and rating, with attendant lack of reliability in measurement. One way in which oral proficiency testing can be assimilated into classroom activities is to require students to participate in a number of speech activities, such as delivering a simulated sales pitch, reading a poem aloud, giving directions on finding a local landmark, and the like. Teachers simply check off for each student the activities that the student has successfully accomplished (Vermont Department of Education, 1978; Pennsylvania Department of Education, 1980). Beyond problems of establishing reliability across different teacher/raters and activities/administrations, type-of-speech checklists may beg the issue of what is to count as a successful speech performance. In order for this type of evaluation to yield meaningful information, each speech activity must be subjected to detailed task analysis and the qualities of mastery-level performance carefully described.

General Impression Marking. Rather than analyzing a performance atomistically, general impression marking assigns a single global score representing a gestalt sense of quality (Cooper, 1977; Godschalk, Swineford & Coffman, 1966). Generally, raters assign general impression marks quickly, based on immediate reactions. Scales usually comprise a five-to-ten interval Lickert type metric ranging from "very poor" to "very good." Evidence suggests that seven-step scales are optimal (Becker & Cronkhite, 1965). With extensive training, raters can render highly reliable ratings on general impression scales. Use of these devices can also be supported by detailed verbal descriptions. (Thompson, 1944) or by videotaped examples (Bowers, 1964) representing typical performance at each level of quality.

Analytic Scales. Analytic scales, widely used in classroom and contest evaluation of speech, consist of more than one item and require raters to judge separate dimensions of speech performance. Each item represents a criterion like "organization," "supporting material," "language fluency," and "voice quality." Each criterion is rated on a multi-interval scale, much as in general impression making. The selection of criteria/items is a major problem of content validity. Becker (1962) suggests that typical analytic scales reflect only three clusters of judgments: content analysis, delivery, and language. As previously mentioned, some analytic scales attempt to incorporate the notion of context-appropriate adaptation by including an item like "speaker adjustment." Frequently, too, analytic scales include a "general merit" item.

Dichotomous Choice Scales. It is generally easier to discriminate the presence or absence of a trait than to discriminate between several qualitative levels of that trait. Consequently, speech rating scales have been developed which ask raters to render forced choice on each of several items (Brooks, 1957). These are essentially a variant of analytic scales wherein each criterion is evaluated by means of a two-interval item. Construction of dichotomous choice scales requires the same analysis of the content/skill domain as does construction of analytic scales.

Rhetorical Trait Scales. Primary trait scoring, introduced by Lloyd Jones (1977) for use in evaluating writing ability, takes as its starting point a rhetorical analysis of the demands peculiar to the communication test situation. The most important, or "primary" rhetorical traits associated with the communication task become items which are scored in accordance with very specific indicators of qualitative levels. Rhetorical trait scales for use in assessing speaking proficiency operate similarly, except that they may include more items or traits, not seeking to distinguish "primary" traits. A separate rhetorical trait scale must be constructed for each assessment situation, thus recognizing the context-dependency of communication competence. The scales convey a high information load, and because of the specificity of quality indicators, they are easily adapted to competency based testing programs. Rubin and Bazzle have developed one rhetorical trait scale (in preparation) for rating performance in a simulated job interview and another for rating performance in a simulated public hearing. One item from the latter scale, by way of illustration, appears as follows:

Conclusion to presentation: (simulated public hearing)

Suggested rating scale -

(1) no conclusion or merely states that remarks are finished
(2) just thanks Commissioners or just restates position
(3) restates position and offers thanks
(4) summarizes or concludes memorably and offers thanks

Content Validity

In discussing speech situations, the paper has already addressed one aspect of content validity: how shall the domain of communication contexts be sampled? Taxonomies of communication roles, functions, and settings abound. One category system that has received repeated attention was expounded by Joos (1961) and includes (1) intimate, (2) casual, (3) consultative, (4) formal and (5) frozen. Testing agencies might consider sampling situations from each of these classes. In any event, it is readily apparent that, despite the ease of rating extended formal discourse, assessment programs which include only public speaking situations have not fully sampled this aspect of the content domain.

Like selection of the speech situation, selection of rating criteria bears on the content validity of speech performance tests. Lists of oral communication competencies have been promulgated (Allen & Brown, 1976; Bassett, Whittington & Staton-Spicer, 1978), as have various conceptual

schemes for analyzing components of communication competence (Wiemann & Backlund, 1980; Wood, 1977). The lack of consensus in identifying qualities inherent in communication competence poses a great barrier to establishing content validity of rating criteria; the content domain remains largely undefined (Larson, 1978).

Surely, however, the parameters of oral communication competence extend beyond the ability to utter well formed language. That is, oral proficiency cannot be equated with general verbal ability, with Standard English syntax and pronunciation, or with "genteel" usage. These factors may contribute to overall speaking ability. But oral proficiency also includes production of nonverbal cues, control over extended patterns of coherence and organization, invention of subject matter, contextually appropriate linguistic and strategic adaptation, anticipation of audience predispositions and ongoing sensitivity to audience responses, management of the mechanics of interaction such as turn-taking, and so forth. Assessment scales which comprise criteria refering exclusively to more narrow language abilities lack content validity.

Reliability

Fair and meaningful performance rating requires consistent judgment and minimal measurement error. In past decades, speech communication researchers devoted substantial attention to reliability of ratings for purposes of research, classroom evaluation, and contest judging. More recent concern about the reliability of performance ratings in large-scale assessment programs has emerged from evaluations of writing ability. In general, research into the reliability of oral assessment is in need of renewed attention.

Intra-Speaker. Students' performances are affected by such factors as time of day, environmental conditions, and physical health—factors extraneous to underlying skill. How many speech samples should be elicited from each individual in order to obtain an accurate estimate of oral proficiency? A single study bearing of this question (Marine, 1965) found a reliability coefficient of about .78 for a single speech performance, increasing to .91 for the reliability of a score based on the average of three performances. Based on this investigation, a single speech sample may be of adequate reliability. In contrast, it is generally deemed advisable to rate at least two writing samples in tests of composition skill (Diederich, 1974).

Equivalent Forms. The nature of the performance task can affect student test scores in two ways. If students are given a choice of topic, then their performance may be affected by difficulties inherent in the subject matter. Rosen (1969), for example, found that in writing examinations students were often "penalized" for choosing certain topics, and that these were topics most often selected by more able students. Rubin and Bazzle (in preparation), however, found no significant differences attributable to choice of topic from among three persuasive propositions in a speech assessment. The issue of equivalent forms reliability is also raised when an oral examination program includes performance tasks representing different speech situations. Is a student's performance in telling a personal narrative, for example, predictive of that same student's score in group discussion? In one testing program, Rubin and Bazzle (in preparation) found that performance

in a simulated job interview correlated .72 with performance ratings derived from a persuasive speech delivered in the context of a simulated public hearing.

Inter-Rater. Speech performance ratings inevitably involve subjective judgments of appropriateness and quality. Moreover, impressions due to extraneous speaker qualities like physical attractiveness may easily contaminate performance ratings. Under poor conditions, rater reliability may range between .50 and .60 (Clevenger, 1963; Marine, 1965). Under proper circumstances it should be possible to achieve inter-rater reliabilities close to .90 (Applebaum, 1974). Since inter-rater reliability is partly a function of the number of judges (Clevenger, 1962), it is always advisable to employ at least two judges for each performance. In one criterion referenced measurement effort, 15 percent of the students were passed by one rater but not by the other (Rubin & Bazzle, in preparation). When cross classifications occurred, a third judge was called in to resolve the certification decision.

Several factors can enhance inter-rater reliability. Among these is the nature of the scales on which ratings are made. Clevenger (1963, 1964) reported that less complex "general merit" items induced the highest reliability, though other researchers dispute this finding (Applebaum, 1974; Thompson, 1944). Reliability is increased when raters are encouraged to utilize all intervals on a scale, including the extreme poles, rather than marking only the intermediate levels of quality (Becker & Cronkhite, 1965).

The conditions under which rating takes place can also affect reliability. Agreement may suffer as a result of rater fatigue. Also, when raters must score a speech after a subsequent student speech has already begun, reliability is not affected, but scores are elevated (Barker, Kibler & Hunter, 1968). Some evidence suggests that raters attain higher consistency when working with videotaped performances in relatively unpressured circumstances, rather than rating "live" performances (Rubin & Bazzle, in preparation).

Raters are, of course, affected by their personality structures (Boatman, 1964), and their ratings of students' oral performances may be influenced by their attitudes towards the speakers' subject matter (Miller, 1964). Judges differ in their leniency, trait, and halo errors, and it may be advisable to employ statistical controls to eliminate these sources of variation (Bock, 1972).

Raters of similar experience and background will naturally respond to speech performances more consistently (Clevenger, 1963). Bowers (1964) found that training of raters affected the absolute values of scores but not reliability. Experience most often demonstrates, however, that rater training can mitigate the effects of rater diversity (e.g., Clevenger, 1963). Effective rater training familiarizes judges with rating criteria and rationales. Ideally, raters should have a hand in the initial formulation of assessment instruments. When raters are exposed to speech samples and share their scores publicly, they can calibrate themselves as a group and come to apply scoring criteria in a consistent and parallel fashion. It is helpful if raters are given the opportunity to periodically recalibrate themselves during the course of rating sessions.

Feasibility

Probably the greatest impediment to adoption of performance testing in general, and speech performance rating in particular, is administrators' concerns about the poor cost-effectiveness of these procedures. Little hard data are available, to either corroborate or allay these concerns. In one comparison of the financial costs of direct and indirect testing of writing ability, it was found that the cost per student of conducting and rating a writing sample was $2.85, while the cost per student of a locally developed standardized multiple choice test was $8.06 (Hudson & Veal, 1980).

Time is a major factor in calculating allocation of institutional resources. The time required of students and regular classroom teachers will vary as a function of the speech situation. Formal speech performances for a class of 20 students may consume no more than a single hour-long class period if students are well prepared. On the other hand, individual interviews may take up to ten minutes per student with an equal amount of interviewer/administrator time required. Rater time must slightly exceed the length of the speech performance. The necessity of multiple raters for each speech performance greatly increases the time needed for oral performance assessment. One program employing two raters for each performance and a "floater" to resolve discrepancies used an estimated .2 person-hours per student for a single formal speech situation (Rubin & Bazzle, in preparation).

Conclusion

Speech performance rating procedures represent a challenge to adequate measurement. They require a substantial allocation of institutional resources. They can, however, provide valid indicators of oral proficiency, and can exert positive effects on classroom instruction. Moreover, if one decides to test speech, there is presently no viable alternative to performance rating techniques. Considering the value to students of instruction in speech communication skills, the commitment to large-scale assessment of oral proficiency via performance rating is, on balance, justifiable.

REFERENCES

Allen, R. R. & Brown, K. L. (Eds.). Developing communication competence in children. Skokie, IL: National Textbook, 1976.

Applebaum, R. L. Inter-rater reliability: A function of scale complexity and rater training? Central States Speech Journal, 1974, 25, 277-281.

Backlund, P. A national survey of elementary and secondary speaking and listening assessment. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, April, 1981.

Barker, L. L., Kibler, R. J. & Hunter, E. C. An empirical study of overlap rating effects. Speech Teacher, 1968, 17, 160-166.

Barnes, D. Situated speech strategies: Aspects of the monitoring of oracy. Educational Research, 1980, 32, 123-131.

Bassett, R. E., Whittington, N. & Staton-Spicer, A. The basics in speaking and listening for high school graduates: What should be assessed? Communication Education, 1978, 27, 293-303.

Bazen, D. The place of conversation tests in oral examinations. English in Education, 1978, 12, 39-50.

Becker, S. L. Rating discussants. Speech Teacher, 1956, 5, 60-65.

Becker, S. L. The rating of speeches; Scale independence. Speech Monographs, 1962, 29, 38-44.

Becker, S. L. & Cronkhite, G. L. Reliability as a function of utilized scale steps. Speech Teacher, 1965, 14, 219-293.

Bock, D. G. Reliability and validity of speech rating scales: Some error effects. Central States Speech Journal, 1972, 23, 145-151.

Boetrom, R. N. Dogmatism, rigidity and rating behavior. Speech Teacher, 1964, 13, 283-287.

Bowers, J. W. Training speech raters with films. Speech Teacher, 1964, 13, 228-231.

Bracewell, R. J., Scardamalia, M. & Bereiter, C. The development of audience awareness in writing. Mimeo, York University, 1978.

Breland, H. M. Can multiple choice tests measure writing skills? College Board Review, 1977, No. 103, 2-7.

Brooks, K. The construction and testing of a forced-choice scale for measuring speaking achievement. Speech Monographs, 1957, 24, 65-73.

Bulloch, A. (Ed.). A language for life. London: Her Majesty's Stationary Office, 1975.

Clevenger, T., Jr. Retest reliabilities of 10 scales of public speaking performance. Central States Speech Journal, 1963, 14, 285-291.

Clevenger, T., Jr. Influence of scale complexity on the reliability of ratings of general effectiveness in public speaking, Speech Monographs, 1964, 31, 153-156.

Cooper, C. R. Holistic evaluation of writing. In C. R. Cooper and L. Odell (Eds.), Evaluating writing. Urbana, IL: National Council of Teachers of English, 1977.

Crowell, L. Rating scales as diagnostic instruments in discussion. Speech Teacher, 1953, 2, 26-32.

Crowhurst, M. On the misinterpretation of syntactic complexity data. English Education, 1979, 11, 91-97.

Diederich, P. B. Measuring growth in English. Urbana, IL: National Council of Teachers of English, 1974.

Educational Testing Service. The English composition test with essay. Princeton, NJ: author, 1979.

Fry, C. L. Training children to communicate to listeners who have varying listener requirements. Journal of Genetic Psychology, 1969, 114, 153-166.

Gary Community Schools Corporation. Oral proficiency program: Secondary schools 1977-1978. Gary, IN: author, 1977-1978.

Gay, G. & Abrahams, R. D. Does the pot melt, boil, or brew? Black children and white assessment procedures. Journal of School Psychology, 1973, 11, 330-340.

Godshalk, F. I., Swineford, F. & Coffman, W. E. The measurement of writing ability. New York: College Entrance Examination Board, 1966.

Hitchman, P. J. Examining oral English in the schools. London: Methuen, 1966.

Hitchman, P. J. CSE tests in spoken English. Educational Research, 1968, 10, 218-225.

Howie-Day, A. M. Metapersuasion: The development of reasoning about persuasive strategies. Unpublished doctoral dissertation, University of Minnesota, 1977.

Hudson, S. & Veal, L. R. Assessing writing competence through writing samples. Report No. 36. Studies in Language Education. Athens, GA: Department of Language Education, University of Georgia, 1979.

Jacobson, R. Linguistics and poetics. In T. Sebeok (Ed.), Style in language. Cambridge, MA: MIT Press, 1960.

Joos, M. The five clocks. New York: Harcourt, Brace and World, 1961.

Keltner, J. (Ed.). Symposium: Problems in the measurement of discussion. Journal of Communication, 1953, 3, 9-33.

Larson, C. E. Problems in assessing functional communication. Communication Education, 1978, 27, 304-309

Lloyd-Jones, R. Primary trait scoring. In C. R. Cooper and L. Odell (Eds.), Evaluating writing. Urbana, IL: National Council of Teachers of English, 1977.

Loban, W. Language development: Kindergarten through grade twelve. Research Report No. 18. Urbana, IL: National Council of Teachers of English, 1976.

Miller, G. R. Agreement and the grounds for it: Persistent problems in speech rating. Speech Teacher, 1964, 13, 257-261.

Marine, D. R. An investigation of intra-speaker reliability. Speech Teacher, 1965, 14, 128-131.

Massachusetts Board of Education. State speaking test (secondary). Boston: author, 1980.

McCaleb, J. Measuring oral communication. English Education, 1979, 11, 41-47.

Mead, N. A. The development of an instrument for assessing functional communication competence of seventeen-year-olds. Unpublished dissertation, University of Denver, 1977.

Mullen, K. A. Direct evaluation of rater and scale in oral interviews. Language Learning, 1978, 28, 301-308.

Palmer, O. Sense or nonsense? The objective testing of English composition. English Journal, 1961, 50, 314-320.

Pennsylvania Department of Education. Speech in the classroom assessment instruments. (E. W. Miller, Ed.). Harrisburg, PA: author, 1980.

Philips, S. V. Acquisition of rules for appropriate speech usage. Monograph Series on Language and Linguistics. No. 23. Washington, DC: Georgetown University Press, 1970.

Rosen, H. An investigation of the effects of differentiated writing assignments on the performance of a selected group of 15/16 year-old pupils. Unpublished dissertation, University of London, 1969.

Rubin, D. L. Psychometric and pedagogical validity in large scale assessment of oral communication skills. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, 1980.

Rubin, D. L. & Bazzle, R. Report of the Glynn County CBE Demonstration Project: Speech Assessment. Brunswick, GA: Glynn County Board of Education, in preparation.

Rubin, D. L. & Piche, G. L. Development in syntactic and strategic aspects of audience adaptation skills in written persuasive communication. Research in the Teaching of English, 1979, 13, 293-316.

Schools Council. The certificate of secondary education - trial examinations - oral English. Examinations Bulletin No. 11. London: Her Majesty's Stationary Office, 1966.

Shantz, C. U. & Wilson, K. F. Training communication skills in young children. Child Development, 1972, 43, 693-698.

Thompson, W. N. An experimental study of the accuracy of typical speech rating techniques. Speech Monographs, 1944, 11, 67-79.

Vermont Department of Education. Basic Competencies: A manual of information and guidelines for teachers and administrators. Montpelier, VT: author, 1977.

Wade, B. Assessing oral abilities at 16+. English in Education, 1978, 12, 52-61.

Wiemann, J. M. & Backlund, P. Current theory and research in communicative competence. Review of Educational Research, 1980, 50, 185-199.

Wilkinson, A. The testing of oracy. In A. Davies (Ed.), Language testing symposium: A psycholinguistic approach. London: Oxford University Press, 1968.

Wood, B. S. (Ed.). Development of functional communication competence: PreK - Grade 6. Falls Church, VA: Speech Communication Association, 1977.